Forecast Combination and Model Selection within a Multi-model Framework: Forecasting Norwegian GDP Growth.

Igor Oleshchuk

Master of Economic Theory and Econometrics

Department of Economics

UNIVERSITY OF OSLO

August 2009

Acknowledgments

This master thesis was written as a part of the Nowcasting project at Norges Bank.

First of all, I want to thank Karsten Gerdrup at the Economics Department for the opportunity to write this thesis. I also want to thank Anne Sofie Jore for providing the dataset that was used during this project and also Leif Anders Thorsrud for all the inspiring discussions about forecasting, as well as other topics. Lastly, the contributions of my supervisor, Christie Smith, have been indispensable and I am truly thankful for the helpful comments and the good advice, that he has offered from beginning to end.

The views, conclusions and any shortcomings represented in this thesis are mine and mine alone.

Igor Oleshchuk

Oslo, August 2009

Summary

The Norwegian central bank currently uses a multi-model strategy to forecast GDP growth. Small individual models are weighted together in SAM (System for Averaging Models) to create a more accurate forecast. The SAM framework also contains 57 bivariate models that uses GDP growth, and one of the indicators that we will use to create models as the endogenous variables. This project adopts a multi-model strategy to cope with uncertainty prevailing about the best strategy for modeling and forecasting economic output, explores the properties of larger VAR models (VARs with three and four variables) and analyzes if they possess the necessary properties to be included in the SAM framework. In other words - if larger-scale VARs can improve on the forecasts made by the bivariate models. This thesis addresses the following questions: Is it limiting to focus on bivariate vector autoregressive models when forecasting GDP growth? Can there be any gain in exploring the three and four variable VAR framework to forecast GDP growth? What combinations of survey indicators perform well in forecasting GDP growth at short horizons? How can one efficiently weight individual forecasting models to deal with model uncertainty?

The results in this thesis clearly indicate that its limiting to focus only on bivariate vector autoregressive models when one uses survey indicators to forecast GDP growth and, there is gain in exploring the three and four variable VAR framework. From the experience gained by performing this exercise one can say that 4VAR models implicitly had the highest posterior weight. Better performing models for short term evaluation of GDP growth than the bivariate have therefore been proposed in this thesis. The combinations of indicators that perform well have mostly been manufacturing indicators (see Table C.1) if one makes judgment from the the *ex post* best models (see Appendix D). From the 10 indicators that were selected by RegSubSets and were used in the weighting exercise, half were manufacturing indicators (the RegSubSets algorithm is explained in Section 3.1 and the selected indicators can be found in Table 3.1). These indicators were selected from the full dataset which contains 57 indicators that are described in Appendix C. One of the indicators (SKI.s - a compound industry indicator) was a part of the 10 selected as well as one of the most frequent regressors in the *ex post* best models, but to create good forecast models according to the RMSE criteria it had to be combined with other survey indicators as well. The indicators “Ressurs” - an indicator for industrial resource scarcity, and “svare” - which indicates change in employment for the construction sector and domestic factories, were the two regressors the exhaustive search algorithm picked most frequently. The algorithm searched for the best possible model (using GDP growth and the survey indicators) adding one indicator at the time. When the best model with one indicator was found, then a search over the models that add one more regressor was

conducted, and the best model with two indicators was recovered. This continued until the best model with GDP growth and 10 indicators was found. The indicator “Ressurs” was a part of all the models that were recovered by the algorithm, hence it was a part of the model containing GDP growth and only one indicator as well as in all the other 9 models that were returned. The indicator “svare” was part of all the models except the one that contained only two variables.

Neither the maximum likelihood based weighted averages nor the mean squared error weighted averages outperformed the simple mean combination of forecasts. One can therefore conclude that in the case of this project, the most efficient way to weight individual forecasting models, to deal with model uncertainty has been a simple mean combination of forecasts.

Contents

List of Figures

Chapter 1

Introduction

Prediction is very difficult, especially if its about the future.
- *Niels Bohr.* - Nobel laureate in Physics.

Predicting the future evolution of gross domestic product (GDP) and inflation is a central concern in macroeconomics. A central bank with a flexible inflation targeting tries to stabilize both prices and aggregate economic activity by adjusting its policy instruments, usually a short term interest rate. To be able to realize these goals, the central bank must understand what the future evolution of these variables is likely to be, and how its actions will affect the future out-turns. While the latter implies that we want to forecast conditional upon a proposed policy path, it is widely accepted that there is considerable inertia in the dynamics of the economy and that monetary policy affects output growth with "long and variable lags," Milton Friedman's often repeated phrase. Given these long lags, even non-structural reduced form models may be useful for forecasting the near-term behavior of the economy. Furthermore, if the policy rules implicit in monetary policy decision making have been relatively stable, then reduced form models may implicitly embody expectational channels, thus nullifying the force of the Lucas critique[1] (Lucas, 1976), which would otherwise suggest that forecasters and policy makers should use structural models to forecast macroeconomic developments.

The Norwegian central bank currently uses a multi-model strategy to forecast GDP growth. Small individual models are weighted together in SAM (System for Averaging Models) to create a more accurate forecast. The SAM framework also contains 59 bivariate models that uses GDP growth, and one of the indicators that we will use to create models as the endogenous variables. This project adopts a multi-model strategy to cope with uncertainty prevailing about the best strategy for modeling and forecasting economic output, explores the properties of larger VAR models (VARs with three and four variables) and analyzes if they possess the necessary properties to be included in the SAM framework. In other words - if larger-scale VARs can improve on the forecasts made

[1]First presented in the article; Econometric policy evaluation: A critique.

by the bivariate models.

This thesis addresses the following questions:

1. Is it limiting to focus on bivariate vector autoregressive models when forecasting GDP growth?
2. Can there be any gain in exploring the three and four variable VAR framework to forecast GDP growth?
3. What combinations of survey indicators perform well in forecasting GDP growth at short horizons?
4. How can one efficiently weight individual forecasting models to deal with model uncertainty?

This paper adopts Bayesian model averaging to deal with model uncertainty and to answer the proposed questions. Bayesian model averaging is an ideal framework for forecast combination, since it provides a rigorous statistical foundation where the weights assigned to the different forecasts arise from posterior probabilities. The models and the combined forecasts have appealing optimality properties given the set of models considered (Min and Zellner, 1993; Madigan and Raftery, 1994).

Besides the Bayesian model averaging method, a variety of different multi-model approaches have been proposed to evaluate the point forecast performance of the different combination schemes. Different model weighting schemes such as equal weights, weights derived from mean squared errors (MSE weights), out-of-sample score-based weights (OSS), Schwartz weights (BIC)[2] and weights from Akaike's information criteria (AIC)[3] are evaluated. The weighted averages are in addition tested in a pseudo out-of-sample forecasting framework, and model inference about the survey indicator performance also drawn.

The rest of the thesis is organized as follows. Chapter two describes the theory of forecasting, model weighting, "ensembling" and forecast combinations that is necessary to understand the results in this thesis. Chapter three presents and discusses the empirical findings, while chapter four proposes topics for future work based on the results of this project and concludes.

[2]See Schwarz (1978).

[3]See, e.g., Akaike (1973, 1974, 1979, 1983) and Burnham and Anderson (2002).

Chapter 2

Theory and framework

2.1 Basics of forecasting

> Forecasting involves the use of information at hand; hunches – formal models, data, etc. – to make statements about the likely course of future events. In technical terms, conditional on what one knows, what can one say about the future? (see Elliott et al., 2006, chap. 1)

Vector autoregressions or VAR modeling is a framework used to capture the evolution of the interdependencies between multiple time series, generalizing the univariate autoregressive (AR) models. VARs started gaining popularity after they were put into use by Sims (1980) to study relationships between the various components that cause fluctuations in business cycles.

In contrast to other fields of science where data could be "exchangeable", i.e. they could be arbitrary reordered without any consequences, in economics we are however often interested in interdependencies between time series. These series may themselves exhibit temporal dependency and we may therefore need to take account for possible lagged relationships. These kinds of relationships between economic time series makes the VAR framework suitable for answering the questions that were proposed in the introduction.

In our analysis, the coefficients of the VARs are estimated by standard OLS estimation method, which is equivalent to maximum likelihood when there are no restrictions on the lagged parameters, see Hamilton (1994). In its most basic form a VAR consists of a set of S endogenous variables $\boldsymbol{Y}_t = (Y_{1t}, Y_{2t}, ..., Y_{St})'$. The VAR($p$) process (where p represents the lag length of the VAR) is then defined as:

$$\boldsymbol{Y}_t = A_0 + A_1\boldsymbol{Y}_{t-1} + \ldots + A_p\boldsymbol{Y}_{t-p} + \boldsymbol{u}_t \tag{2.1}$$

where the A_i's are $(S \times S)$ coefficient matrices for $i = 0, 1, ..., p$ and $\boldsymbol{u}_t = (u_{1t}, u_{2t}, ..., u_{St})'$. The VAR assumptions are the same as for a general time series regression. Expected values of the error terms conditional on the regressors are equal to 0, the variables in the different time series become independent of each other when the lag length is large, all variables in the system of equations have nonzero, finite fourth moments and there is no perfect multicollinearity.

We are interested in forecasting the value of a variable Y_{t+h} (where h represents the forecast horizon) based on a set of variables $X_t, X_{t-1}, X_{t-2}, ..., X_1$ which are observed at date t. We can denote $Y^*_{t+h|t}$ as the forecast of Y_{t+h} based on $X_t, X_{t-1}, X_{t-2}, ..., X_1$. To evaluate the usefulness of the forecast, we need to specify a loss function. This loss function represents the "disutility" that arises when the forecast is off target by a particular amount. The standard approach is to specify a quadratic loss function where we choose the forecast $Y^*_{t+h|t}$ such that it minimizes:

$$E(Y_{t+h} - Y^*_{t+h|t})^2 \tag{2.2}$$

This expression is known as the mean squared error (*MSE*) and is denoted:

$$MSE(Y^*_{t+h|t}) \equiv E(Y_{t+h} - Y^*_{t+h|t})^2 \tag{2.3}$$

The forecast with the smallest *MSE* turns out to be the expectation of $Y^*_{t+h|t}$ conditional on $X_t, X_{t-1}, X_{t-2}, ..., X_1$[1]:

$$Y^*_{t+h|t} = E(Y_{t+h}|\boldsymbol{X}_t) \tag{2.4}$$

In this particular exercise the *RMSE* will be used instead of the *MSE* (where $RMSE = \sqrt{MSE}$), to evaluate the point forecasts from different models and various combinations of models. The $RMSE$ results in the same ordering as the MSE and has the same scale as the underlying variable being forecasted.

2.2 Bayesian Approach

To motivate one of the weighting approaches used in this thesis (BIC weighting scheme), and to some extent the model ensembles as well, I will devote a little space to address the Bayesian approach to model inference.

The Bayesian approach to inference, decision making and forecasting, involves conditioning on what is known to make statements about something that is unknown. The parameters of the model are no more known than future values of the data time series that are going to be generated by the model and the Bayesian approach treats these two types of unknowns in symmetric fashion (e.g., Elliott et al., 2006). The future values of

[1]The proof of this claim can be found in Hamilton (1994, page 73).

an econometric time series just constitute another function of interest for the Bayesian analysis.

Conditioning on what is known means that we need to use prior knowledge of the structures and reasonable parameterizations. This key feature of the Bayesian approach can be viewed as an implication of the principle that one should fully specify what is known and what is unknown, and after that condition on the known and specified information to make probabilistic statements about the unknown information.

Bayesian model averaging (BMA) can be thought of as a Bayesian approach to forecasting. The idea of BMA is to take forecasts from many different models, and to assume that one of them is the true model, but that the researcher does not know which one this is. The weights of the individual forecasts are computed as posterior probabilities that the models are correct. The individual forecasts in the BMA framework are model-based and are the posterior means of the variable to be forecast, conditional on the selected model. BMA first started out as a subject of substantial research outside the field of economics, mainly in the area of computational sociology (Hoeting et al., 1999), but has also been applied in forecasting of economic variables (e.g., Wright, 2008, 2009).

In standard Bayesian analysis, parameters of a given model are treated as random and are distributed according to a prior distribution. In BMA the binary variable that indicates the probability whether a given model is true is also treated as random and distributed according to a prior distribution. These variables will be estimated by BIC values for the individual models in the next chapter.

2.3 Model space

This section describes the models that were developed to perform the pseudo out-of-sample forecasting exercises. This is a "pseudo" out-of-sample forecasting exercise - because the final vintage of the data set is used (in place of vintages of real time data), since a proper real-time data set for Norwegian GDP is not yet available. It is an "out-of-sample" forecasting exercise because only the data available up to time t are used to estimate parameters and forecast outturns for true $t+1, ..., t+5$. This exercise is repeated for various t. The forecast mimics the forecasting process that would occur if the forecast was done in real time (without having the benefit of knowing future values of the variable). The VAR models listed and described in the paragraphs below are used to provide out-of-sample forecasts, which are evaluated with the RMSEs. In the out-of-sample model framework the first forecasted value is Q1 2000, hence the first vintage uses the data up to Q4 1999. The last vintage uses the data up to Q3 2008 to forecast Q4 2008 to Q4 2009. There are

36 vintages of data, and hence 36 "forecast rounds", and the maximum forecast horizon is $h = 5$ quarters.

One of the main goals of this project is to find the best possible models to forecast GDP growth in Norway at short horizons, based on the available set of possible indicators. The models in the model space are used to provide out-of-sample forecasts for GDP growth 1 to 5 steps ahead, where each step represents a quarter. The VARs are built up from previous values of GDP-growth and indicators that are correlated with it. Identifying the indicators that can help to forecast GDP growth in the best possible manner is therefore a central part of the thesis.

To build the models that are used in this project the computational language R was used. R is a variant of the S language, which was originally developed at Bell laboratories by Rick Becker, John Chambers and Allan Wills. R was chosen to create the models in this project because R is a free license language and contains packages that have been developed to address multi-model inference. Packages like the VAR package (Pfaff, 2008) have simplified the programming part of this project to some extent and have made it possible to do a serious amount of number crunching. This analysis has identified the best single (out-of-sample) forecasting models within a multivariate VAR framework. These single models are used as a reference point to evaluate the performance of the statistically significant models in which model uncertainty has been accounted for.

2.3.1 Benchmark models

The first benchmark models that were developed were two simple autoregressive (AR) models where GDP growth was forecasted based on previous values following (Marcellino, 2008) (iterated forecast method with a rolling and a recursive window, where the rolling window contains 28 observations while the start date for the recursive window is Q1 1993). The AR models were made to establish whether the AR models should be used to forecast GDP, or whether they should be substituted by more sophisticated specifications. Univariate linear models are often more robust than their multivariate counterparts (see, for example, (Banerjee and Marcellino, 2006)) and are thus a common benchmark used to assess more complex forecasting methods. The maximum lag length of the AR models was set to 4 lags.[2]

[2]This is based on the results of the *Marcellino, Stock and Watson* paper (Marcellino et al., 2006) A more precise description of the empirical research that supports this conclusion can be found in the article. This maximum lag length was also used for the rest of the models in the model space since the time series used in this project are rather short. Quarterly GDP growth being the longest series where the first observation is the second quarter of 1979. Some of the indicator time series start as late as the first quarter of 1993.

After the two AR models were developed, 114 bivariate models were made using the iterated forecasting method (57 with the rolling and 57 with a recursive window[3]), where each model contains GDP growth plus one of the indicators. The iterated forecasting method will be described in Section 2.6. These models were then tested with 4 to 1 lags, and compared to Norges Bank's models, since they already have all these bivariate[4] models in their model framework, to verify that the modeling setup was correct and that there were no programming errors.

Every model that was made was modified into two different editions. As an example, out of the 2VAR model was made, it was split up into following editions:

Model_name	Window	Forecast method
2VAR	Recursive	Iterated
2VAR	Rolling	Iterated

Table 2.1: Model editions. All models can be run with lag length 1 - 4.

After the 2VAR models were created, 3VAR models of the same structure were made. These models create creates all possible combinations of models with GDP growth and two indicators. (There were 57 indicators to choose from) The total amount 3VAR models made by each edition of the big looping 3VAR model becomes (for a fixed lag length):

$$\left[\frac{57!}{2!55!} = 1596\right]$$

The last models that were created were the big 4VAR models with 2 editions (rolling and recursive) by the same framework as the 3VAR models, only the dataset was shortened from 57 to the 25 best indicators after looking at the performance of the 3VARs. This is because running the 4VAR over all possible combinations resulted in a model framework of 39260 models. This was first attempted, but since this took more than 36 hours to run a big looping 4VAR model, and had to be done for different lag lengths, the dataset was shortened to 25 indicators. The computational time for each 4VAR model edition was now down to 4 hours, which made it more feasible computationally. Each edition of

[3]Combines GDP growth with all the indicators. A recursive window increases with the amount of vintages. When one more vintage is simulated then one more observation is added at the end of the data set, hence the size of the recursive window increases with the amount of vintages. The size of the rolling estimation window stays constant for the whole out-of-sample simulation period, and is from Q1 1993 to Q4 1999 in this thesis (28 observations).

[4]At Norges Bank BiVAR was the original terminology for the bivariate VARs, but with a Norwegian accent it sounded too much like BVAR and therefore the bivariate VARS became known as 2VARs. In this thesis trivariate are referred to as 3VARs, and VARs with four variables are known as 4VARs. The term 2VAR should not be confused with VAR(2) which represents a VAR with 2 lags.

the 4VAR now created 2300 new models that were added to the model framework. After this the best models were selected by the RMSE criteria.[5] A possible exploration of the 5VAR models was considered to take too much time in computational terms.

The parameter estimates of these models at time t reflect the information or data that were available at time t. However, the “best” models that we report depend on knowledge of the entire sample. A genuine selection strategy would involve picking the “best” model at time t and, this “best” model would likely change over time as new data arrives. That is, these models are subject to selection bias (Miller, 2002, p.7). The data used to estimate the parameters are the same data used to select the regressors in the first place. In the rest of this thesis we will concentrate on combinations of models.

2.3.2 Weighted averages

The models described in this subsection are constructed to answer the proposed questions for this thesis. The models described in section 2.3.1 are only used as benchmark models since they are subject to selection bias.

To draw valid model inference the 2VAR, 3VAR and 4VAR models were ensembled[6] into 3 different model classes and weighted with different weighting schemes. The weighting schemes that were used are Akaike information criteria (AIC), Bayesian information criteria (BIC), Out-of-sample score-based (OSS), Mean squared error (MSE) and equal weights. Furthermore, the original dataset was reduced from 57 to 10 indicators with the RegSubSet method (Miller, 2002), but more on this subject in the next chapter.

2.4 Dataset

The data that are used in this project are mainly from the “Market state barometer” (Konjunkturbarometeret in Norwegian) developed by the Norwegian Bureau of Statistics, (SSB). The “Market state barometer” is a qualitative survey that records the beliefs of enterprise leaders about production, the economy’s capacity usage, order reserves and supply. A closer description of the indicators that are used for forecasting can be found in the appendix (SSB, 2008).[7] Survey indicators are believed to be important for forecasting future developments in in GDP growth (Banerjee et al., 2003; Dovern, 2006; Dresse and Nieuwenhuyze, 2008). Some of the indicators (such as “Employment” and “Other”) are

[5]See Appendix for the full list of models in the model space.

[6]The term “ensembled forecasts” descends from meteorological literature.

[7]A much richer description of the Market state barometer can be found at http://www.ssb.no/emner/08/05/10/kbar.

not a part of the “Market state barometer”, but are still SSB data.[8]

2.5 Out-of-sample forecasting

The forecasting is performed in a pseudo out-of-sample framework. Pseudo out-of-sample gives a sense of how well the model has been forecasting at the end of the sample when the actual forecast is performed. The steps of the exercise are done in the following way:

1. Let $T_0 = Q1, 1993$ denote the first observation used in estimation of the regressions,[9] $T_1 = Q1, 2000$ denote the date at which the first pseudo out-of-sample forecast is made, and $T_2 = Q3, 2008$ the date at which the final pseudo out-of-sample forecast is made (last observation in the data set). Here $\frac{T_1-T_0}{T_2-T_0}$ is fixed with respect to the sample size following Rissanen (1986) and Wei (1992) who also use a fixed $\frac{T_1-T_0}{T_2-T_0}$.

2. We then estimate the forecasting regression using the shortened data set for $t = T_0, ..., (T_1 - 1)$. The regression is a standard OLS estimation.

3. Then the forecasts for the 1st to 5th step are computed for this shortened sample: $Y^*_{T_1|T_1-1}$ to $Y^*_{T_1+4|T_1-1}$.

4. Now the forecast error is computed for the 5 forecasts: $Y_{T_1} - Y^*_{T_1|T_1-1}$ to $Y_{T_1+4} - Y^*_{T_1+4|T_1-1}$ and from this the Bias and RMSE.

5. After this regression step $2-4$ are repeated for the remaining dates, T_1 to $T_2 - 1$. The regression is re-estimated at each date, on a new pseudo vintage of data. The exercise is performed with both recursive and rolling estimation windows.

The final forecast date depends on the forecast horizon, and is the last available observation plus the forecast horizon h. The pseudo out-of-sample forecast error is $e_{t+h|t} = Y^*_{t+h|t} - Y_{t+h}$, where $Y^*_{t+h|t}$ is the pseudo out-of-sample forecast.

To calculate the forecast bias and the RMSE in an out-of-sample framework, the following formulas are used:

$$Bias(h) = \frac{1}{T_2 - T_1 + (2-h)} \sum_{t=T_1-1}^{T_2-h} e_{t+h|t}, \text{ for } h = 1, 2, ..., 5 \tag{2.5}$$

$$RMSE(h) = \left(\frac{1}{T_2 - T_1 + (2-h)} \sum_{t=T_1-1}^{T_2-h} e^2_{t+h|t} \right)^{\frac{1}{2}}, \text{ for } h = 1, 2, ..., 5 \tag{2.6}$$

[8] These indicators have been transformed to fit the purposes of forecasting by Anne Sofie Jore from Norges Bank.

[9] In this project T_0 is Q1 1993 because the first observations of the indicators used in the VARs with the shortest sample size start at that date.

The size of the denominator decreases as h increases because all final pseudo out-of-sample forecasts that can be compared with actual values ($\forall h$) are made at T_2 since it is our last observation for GDP growth in the data set. The part of the data set that is used for Bias and RMSE calculation is therefore shorter for $h = 5$ than for $h = 1$ since the first pseudo out-of-sample forecast that is made for $h = 1$ is T_1 and $T_1 + 4$ for $h = 5$.

Since this is a VAR framework we also automatically forecast the future values of the indicators in all vintages, but GDP growth is the only variable that is of interest when it comes to forecasting in this project, therefore the Bias and RMSE are calculated in the same way as they would have been in a simple AR model.

2.6 Iterated forecasting method

This subsection offers a brief explanation of the iterated forecasting method. Suppose that we are working with vintage 1 and want to forecast 5 periods ahead (which is max h for this project) in a VAR framework containing S endogenous variables, with the lag length equaling p and $\widehat{A}_i$'s denoting the $(S \times S)$ estimated coefficient matrices for $i = 0, 1, ..., p$. Using the iterated forecasting method to forecast one period ahead, the following model is used:

$$\boldsymbol{Y}^*_{T_1|T_1-1} = \widehat{A}_0 + \widehat{A}_1\boldsymbol{Y}_{T_1-1} + \widehat{A}_2\boldsymbol{Y}_{T_1-2} + \ldots + \widehat{A}_p\boldsymbol{Y}_{T_1-p} \tag{2.7}$$

where $\boldsymbol{Y}_{T_1-p}$ is a vector containing the values of S endogenous variables at $t = T_1 - p$. If we want to make an iterated forecast 2 periods ahead:

$$\boldsymbol{Y}^*_{T_1+1|T_1-1} = \widehat{A}_0 + \widehat{A}_1\boldsymbol{Y}^*_{T_1|T_1-1} + \widehat{A}_2\boldsymbol{Y}_{T_1-1} + \ldots + \widehat{A}_p\boldsymbol{Y}_{T_1-p+1}.$$

The same structure follows for period 3 and 4. The iterated forecast 5 periods ahead then follows:

$$\boldsymbol{Y}^*_{T_1+4|T_1-1} = \widehat{A}_0 + \widehat{A}_{1|}\boldsymbol{Y}^*_{T_1+3|T_1-1} + \widehat{A}_{2|t+5}\boldsymbol{Y}^*_{T_1+2|T_1-1} + \ldots + \widehat{A}_p\boldsymbol{Y}_{T_1-p+4}. \tag{2.8}$$

When all 5 forecasts are done, the RMSE is calculated for each of the forecasts, another observation of the real data is added (or one is added and 1 is removed from the tail of the data sample if we are working with a rolling window) to make vintage number 2, and this continues until we reach the end of the data sample.

2.7 Forecast combinations

Combining point forecasts to improve forecast performance goes back at least to the work of Bates and Granger (1969) who combined two separate sets of forecasts of airline passenger data to form a composite set of forecasts. Combination has gained popularity since forecasters want to draw valid model inferences and discarded forecasts nearly always contain some useful independent information. The independent information can be of two kinds. First, a discarded forecast can be based on variables that are not featured in the chosen forecast model. Second, the discarded forecast makes a different assumption about the form of the relationship between the variables (Bates and Granger, 1969). The second case in particular does not necessarily lead to a situation in which a combined forecast improves upon the better individual forecast.

The formulas for the different weights are displayed in Table 2.2 and the next 4 subsections describe the different weighting schemes that are used to make point forecast combinations in this project.

Equal weights (EW) $\Longrightarrow \omega^{(i)}_{t+h\mid t} = \frac{1}{N} \ \forall \ i,t,h.$
Mean squared error weights (MSEW) $\Longrightarrow \omega^{(i)}_{t+h\mid t} = \frac{1/MSE^{(i)}_{t+h\mid t}}{\sum_{j=1}^{N} 1/MSE^{(j)}_{t+h\mid t}} \ \forall \ t,h.$
Akaike information criteria weights (AICW) $\Longrightarrow \omega^{(i)}_{t} = \frac{exp\Big\{-\frac{1}{2}\Delta^{(i)}_t(AIC)\Big\}}{\sum_{j=1}^{N} exp\Big\{-\frac{1}{2}\Delta^{(j)}_t(AIC)\Big\}} \ \forall \ t.$
Bayesian information criteria weights (BICW) $\Longrightarrow \omega^{(i)}_{t} = \frac{exp\Big\{-\frac{1}{2}\Delta^{(i)}_t(BIC)\Big\}}{\sum_{j=1}^{N} exp\Big\{-\frac{1}{2}\Delta^{(j)}_t(BIC)\Big\}} \ \forall \ t.$
Out-of-sample score-based weights (OSSW) $\Longrightarrow \omega^{(i)}_{t+h\mid t} = \frac{exp\Big(\Lambda^{(i)}_{t+h\mid t}\Big)}{\sum_{j=1}^{N} exp\Big(\Lambda^{(j)}_{t+h\mid t}\Big)} \ \forall \ t,h.$

Table 2.2: Weight formulas used in this project. The following sub chapters explain the notation used in the table.

2.7.1 Equal weights (EW)

Equal weights were introduced by Bates and Granger (1969) and are the simplest forecast combination method. Although combined forecasts where the individual forecasts are given equal weights are acceptable for illustrative purposes, one wishes to give greater weights to the set of forecasts that contain the lower mean squared errors. On the other hand, equal weights for point forecast combination have often proven to be better than more sophisticated weighting schemes when the robustness of the forecast combinations is tested pseudo out-of sample (Stock and Watson, 2004). Formally, $\omega_{t+h|t}^{(i)}(EW) = 1/N$ $\forall\ t, h$ and i. Where i denotes an individual model, N the number of models and h the forecast horizon which is $1, \dots, 5$ in this project.

In this project, equal weights are used in two different ways. One approach is labeled equal weight for classes (EWC) where individual models are first given equal weights within their model class (2VAR, 3VAR and 4VAR model class) and then each of the classes are given equal weights when they are ensembled together. The second approach is equal weighted models (EWM) which assigns equal weights to all models that are ensembled together independent of model class.

2.7.2 Mean squared error weights (MSEW)

The concept of ranking models with respect to their relative mean squared prediction error performances computed over a window of previous observations was also introduced by Bates and Granger (1969). In this project we calculate the MSE weights following Kascha and Ravazzolo (2008). Although their paper focuses on density forecasts, the calculation of MSE weights for point forecasts is quite similar.

These MSE weights are not optimal in a linear framework since MSE weights ignore the correlation structure between forecasts (Granger and Ramanathan, 1984), but tend to outperform the more sophisticated weighting schemes since the correlation matrix of the forecast errors is quite difficult to estimate.

Weights derived from the relative inverse mean squared error of an individual model i for the h-step ahead point forecast take the following form:

$$\omega_{t+h|t}^{(i)} = \frac{1/MSE_{t+h|t}^{(i)}}{\sum_{j=1}^{N} 1/MSE_{t+h|t}^{(j)}}, \tag{2.9}$$

where

$$MSE_{t+h|t}^{(i)}(Y_{t+h|t}^{*}) = \frac{1}{t-h-\underline{t}+1} \sum_{\tau=\underline{t}}^{t-h} \left(Y_{\tau+h} - Y_{\tau+h|\tau}^{*(i)}\right)^2, \tag{2.10}$$

and

$$\sum_{i=1}^{N} \omega_{t+h|t}^{(i)} = 1, \forall t, h.$$

Letter $\underline{t}$ is the beginning of the evaluation period minus 1 while h, t, i and N have the same interpretation as for equal weights. In this project $\underline{t}$ is last observation in the dataset before the the first pseudo out-of-sample forecast for step h is made (for the first vintage, step 1, $\underline{t}$ is Q4 1999).

2.7.3 Akaike information criteria weights (AICW)

Akaike weights are derived from the Akaike information criteria (AIC) which Hirotsugu Akaike proposed in his seminal paper (Akaike, 1973). The main idea of this information criteria is to use the Kullback-Leibler (K-L) information (distance) as a fundamental basis for model selection.[10] The K-L distance can not however, be computed without full knowledge about the data generating process and the parameters of all the candidate models, which is never the case when we forecast economic variables. To overcome this problem, Akaike proposed an information criteria that estimated the K-L information, based on a model's maximized empirical log-likelihood function.

To calculate the Akaike weights I need to calculate the AIC values for the individual models. To do that I must define the empirical log likelihood (Λ_i) at its maximum point for model i.[11] Formally:

$$\Lambda_i = \left(\frac{T_i - p_i}{2}\right) \left(log|\widehat{\mathbf{\Omega}}_i^{-1}| - nlog(2\pi) - n\right) \tag{2.11}$$

where

$$\widehat{\mathbf{\Omega}}_i \equiv \begin{bmatrix} var\left(\widehat{\epsilon}_1^i\right) & cov\left(\widehat{\epsilon}_1^i, \widehat{\epsilon}_2^i\right) & ... & cov\left(\widehat{\epsilon}_1^i, \widehat{\epsilon}_n^i\right) \\ cov\left(\widehat{\epsilon}_1^i, \widehat{\epsilon}_2^i\right) & var\left(\widehat{\epsilon}_2^i\right) & ... & . \\ . & cov\left(\widehat{\epsilon}_2^i, \widehat{\epsilon}_3^i\right) & ... & . \\ . & . & ... & cov\left(\widehat{\epsilon}_{n-1}^i, \widehat{\epsilon}_n^i\right) \\ cov\left(\widehat{\epsilon}_1^i, \widehat{\epsilon}_n^i\right) & . & ... & var\left(\widehat{\epsilon}_n^i\right) \end{bmatrix}$$

AIC_i then becomes:

$$AIC_i = -2(\Lambda_i) + 2K_i \tag{2.12}$$

with $\widehat{\mathbf{\Omega}}_i$ being the variance - covariance matrix for model i, $T_i - p_i$ being the length of the original data sample minus the lag length of the VAR, n represents the number of variables in the system and K_i is the number of parameters estimated in the system including the intercepts and the variances - covariances. So instead of estimating

[10]For more information on K-L information see Burnham and Anderson (2002, chap. 2).

[11]For more details see Hamilton (1994) page 295-298.

the K-L distance between two models, one estimates the expected distance between the fitted model and the unknown true mechanism that actually generated the observed data.

To calculate the weight for model i it is common to compute, for each model, the differences in AIC with respect to the AIC of the best candidate model (Akaike, 1978) (Burnham and Anderson, 2002):

$$\Delta_i(AIC) = [AIC_i - min(AIC)]\,,$$

where min(AIC) is the smallest value of AIC in the model set. The best model is the one with lowest AIC value / highest maximized log likelihood. Its important to notice that these differences are calculated for models from the same model class, vintage and forecasting step h. So the equation above can be written as:

$$\Delta^{(i)}_{t+h|t}(AIC) = \left[AIC^{(i)}_{t+h|t} - min(AIC)_{t+h|t}\right],$$

if one wants to be more precise.

Akaike weights derived for an individual model (i), at time t and h-step ahead point forecast take the following form:

$$\omega^{(i)}_{t+h|t} = \frac{exp\left\{-\frac{1}{2}\Delta^{(i)}_{t+h|t}(AIC)\right\}}{\sum_{j=1}^{N} exp\left\{-\frac{1}{2}\Delta^{(j)}_{t+h|t}(AIC)\right\}} \tag{2.13}$$

$$\sum_{i=1}^{N}\omega^{(i)}_{t} = 1, \forall t, h.$$

Where N is the number of models and h, t, i have the same definitions as before.

The AIC weights above are written in a general form, but are used as in-sample weights in this project (based on the in-sample performance of the models). The AIC weights are therefore not re-estimated for every forecast step, but estimated for every vintage for each individual model and multiplied with all of the out-of-sample point forecast that the individual model i creates, so all forecast steps h are assigned the same weight within a vintage. This is also the case for BIC weights. The in-sample AIC weight for the individual model i can be written as:

$$\omega^{i}_{t}(AIC) = \frac{exp\left\{-\frac{1}{2}\Delta^{i}_{t}(AIC)\right\}}{\sum_{j=1}^{N} exp\left\{-\frac{1}{2}\Delta^{j}_{t}(AIC)\right\}}. \tag{2.14}$$

Although the AIC's are used to weigh models instead of using it to select the lag length of the VAR's and hence the choose the most parsimonious models,[12] an AIC weighting

[12]The *principle of parsimony* was formalized by Box and Jenkins in 1970 and is a bias versus variance tradeoff. In general, bias decreases and variance increases when the dimension of a model increases. The principle therefore tells us to select the smallest possible number of parameters which still can give an adequate representation of the data (Burnham and Anderson, 2002).

scheme may still have some advantages over a weighting scheme which uses the equal model weights since it assigns the heaviest weight to the individual model within a class that has the highest maximized log likelihood (AIC_{min}) if the models have the same number of parameters.

2.7.4 Bayesian information criteria weights (BICW)

To calculate Bayesian Information Criteria (BIC) weights (Schwarz, 1978) Λ_i is calculated in the same manner as for the AIC, but the penalty term for the BIC is different.

$$BIC_i = -2(\Lambda_i) + K_i log(T_i - p_i) \tag{2.15}$$

and

$$\Delta_i(BIC) = [BIC_i - min(BIC)]$$

Just as raw AIC values may be converted to Akaike weights, raw BIC values can be converted to BIC model weights (or Schwarz weights). Schwarz weights can be obtained by replacing the AIC with BIC in the ω_i equation above:

$$\omega_t^i(BIC) = \frac{exp\left\{-\frac{1}{2}\Delta_t^i(BIC)\right\}}{\sum_{j=1}^{N} exp\left\{-\frac{1}{2}\Delta_t^j(BIC)\right\}} \tag{2.16}$$

This can be done because I set equal 'reference' prior probabilities on the individual models in each class where prior probability for model i is set to equal $1/N$ for all models where N is still the number of models in each model class.

Although the equations of AIC and BIC look very similar, they originate from quite different frameworks. The BIC assumes that the true model for forecasting GDP growth is in the set of candidate models, and it measures the degree of belief that a certain model is the true data-generating model. The AIC does not assume that any of the candidate models is necessarily true, but rather estimates for each model the Kullback Leibler discrepancy, which is a measure of distance between the probability density generated by the model and reality. The Bayesian information criteria also favors simple models (i.e., those with fewer parameters) to a greater extent than AIC favors.

2.7.5 Out-of-sample score-based weights (OSSW)

In this project Out-of-sample score-based weights (OSSW) equal the AICW when they are computed out-of-sample and re-estimated for different forecast steps and vintages when the models have the same number of parameters (see Appendix A.1 for the details). The weight for the individual model i becomes

$$\omega_{t+h|t}^{(i)} = \frac{exp\left(\Lambda_{t+h|t}^{(i)}\right)}{\sum_{j=1}^{N} exp\left(\Lambda_{t+h|t}^{(j)}\right)}. \tag{2.17}$$

The estimation of OSS weights in this project is quite similar to the work of Eklund and Karlsson (2007) who find that the forecast combinations weighted using predictive likelihoods have good small sample properties.

2.8 Model ensembles

This section explains how the models are ensembled together to test if 3VAR and 4VAR models can improve on the 2VAR framework already used by Norges Bank.

To avoid discrimination against any of the model classes when they are ensembled together, and because there is no *ex ante* knowledge about how the different model classes will perform, the same prior probabilities are assigned to each model class. All models in each of the classes[13] are used when the analysis is performed (Chipman et al., 2000) so the prior probabilities on the individual 2VAR models will be higher in the ensembled weighting scheme since there are fewer of them compared to 3VAR and 4VAR models, while the posterior probabilities also depend on the likelihood of the individual models.

The regressors that are chosen to create the VAR models that will be used to make the weighted model averages are picked by an exhaustive search algorithm called RegSubSets (this procedure will be explained in more detail in the next chapter) and count 10 regressors. These regressors are described in Table 3.1. When the 10 best regressors are chosen, we have 10 2VAR models, $(10!)/(2!8!) = 45$ 3VAR and $(10!)/(3!7!) = 120$ 4VAR models; to make notation easier, $N_2 = 10$, $N_3 = 45$ and $N_4 = 120$. As a result, the individual models weighted together by EWC, MSEW and OSSW follow the weighting scheme described below when the different model classes are ensembled together into one averaged point forecast:

$$\sum_{i=1}^{N_2} \omega_{t+h|t}^{(i)} = 1,$$

if one only looks at the performance 2VARs weighted together,

$$\frac{1}{2}\left(\sum_{i=1}^{N_2} \omega_{t+h|t}^{(i)} + \sum_{i=1}^{N_3} \omega_{t+h|t}^{(i)}\right) = 1$$

if one looks at the performance 2VAR and 3VAR models weighted together,

$$\frac{1}{3}\left(\sum_{i=1}^{N_2} \omega_{t+h|t}^{(i)} + \sum_{i=1}^{N_3} \omega_{t+h|t}^{(i)} + \sum_{i=1}^{N_4} \omega_{t+h|t}^{(i)}\right) = 1$$

[13]There are 3 model classes: All possible 2VARs, 3VARs and 4VARs.

if one looks at the performance 2VAR, 3VAR and 4VAR models weighted together, and finally if one wants to evaluate M different model classes the general formula becomes

$$\frac{1}{M}\sum_{j=1}^{M}\sum_{i=1}^{N_j}\omega_{t+h|t}^{(i)} = 1 \tag{2.18}$$

the $\frac{1}{M}\omega_{t+h|t}^{(i)}$ (posterior probabilities) are then multiplied with the individual forecasts and summed to create a single point forecast in the simulated out-of-sample exercise, with $M \in [1, 3]$ for this project. Furthermore, the equations hold if the model weights are calculated by equal weights for classes , mean squared error or out-of-sample score-based approach.

When, on the other hand the model weights are calculated by Akaike (AIC), Bayesian (BIC) or equal weights for all models (EWM) the esembling scheme becomes a little different. When we deal with "grand ensembles" of all 3 model classes to calculate an averaged point forecast the weight attached to a point forecast made by an individual model i at time t becomes:

$$\omega_{AIC_t}^{(i)} = \frac{exp\left\{-\frac{1}{2}AIC_t^i\right\}}{\sum_j^{N_2} exp\left\{-\frac{1}{2}AIC_t^j\right\} + \sum_j^{N_3} exp\left\{-\frac{1}{2}AIC_t^j\right\} + \sum_j^{N_4} exp\left\{-\frac{1}{2}AIC_t^j\right\}}$$

for Akaike weights and

$$\omega_{BIC_t}^{(i)} = \frac{exp\left\{-\frac{1}{2}BIC_t^i\right\}}{\sum_j^{N_2} exp\left\{-\frac{1}{2}BIC_t^j\right\} + \sum_j^{N_3} exp\left\{-\frac{1}{2}BIC_t^j\right\} + \sum_j^{N_4} exp\left\{-\frac{1}{2}BIC_t^j\right\}}$$

for Bayesian weights with

$$\sum_{i=1}^{N}\omega_t^{(i)} = 1$$

for both AIC and BIC weights and with $N = N_2 + N_3 + N_4$. The reason for this new ensembling scheme which gives all models equal prior probabilities independent of their model class is because if AIC and BIC weights are derived within specific model classes before the models are ensembled, then the weights become equal and also equal OSS weights (the reason for this is explained in Appendix A.1). Furthermore, when the ensemble with equal weights for all models (EWM) is used, all models are assigned the weight $\frac{1}{N}$ independent of model class.

One negative side with using weighting schemes that assign equal prior probabilities to individual models is that the model class that contains the most models (4VAR) is given the highest prior weight. By this we indirectly say that 4VAR models matter more than 2VAR and 3VAR models, since the prior probability on 4VARs is greater. 4VARs will therefore get a head start in answering the third question in the introduction of this thesis.

“How limiting is it to focus on two variable vector autoregressive models when forecasting GDP growth?”

Another important point is that the difference between in sample AIC and BIC weights (within the framework of this project) is negligible even if the weights are calculated based on in-sample model performance and models of different size are ensembled together before the weights are calculated. The difference between the averaged point forecasts created by ensembled 2VAR models, “grand ensembles” of 2VAR and 3VAR models, and 2VAR, 3VAR and 4VAR models also becomes negligible. This is a quite unusual result and reasons for this are carefully explained in Appendix A.2.

An ensemble framework like the first one described above is convenient for analyzing whether adding more similar models actually improves forecasting performance and at the same time gives an indication about which model size can be considered as over parametrized.[14] These topics are analyzed in the next chapter.

[14]Following the *principle of parsimony* it will be the model class that gives no improvement in performance when its models are added to the ensemble with models of smaller size.

Chapter 3

Empirical results

The main results are stated and discussed in this chapter.

3.1 Regression subsets

To answer the questions that were proposed in the beginning of this thesis, the dataset of 57 survey indicators had to be reduced. The number of regressors that this analysis is built on is 10 because of computational limitations (averages with more than 10 indicators would result in more than 120 4VAR models and the computational time for the weighting schemes would be too long). To choose these 10 indicators, an exhaustive search algorithm was used (Miller, 2002).

The exhaustive search method that was used comes from the “regsubsets” package implemented with R. The algorithm searches for the best possible model adding one indicator at the time where best means the one with the lowest possible BIC value or highest adjusted R^2 . When the best model with one indicator is found, then a search over the models that add one more regressor is conducted, and the best model with two indicators is recovered. This continues until the best model with 10 indicators is found (see graph of the process on Figure 3.1 and 3.2, the second figure displays the adjusted R^2 for the selected models).

One problem with using this method is that it is an in-sample algorithm, it does not explicitly specify the models, only which variables it includes and not their preferred lag length. To do that one would have to shorten the column of each variable the same amount of times as the number of lags one wants to investigate and include the new columns in the dataset. So to investigate 4 lags of the 57 variables, one would need to include 57×4 columns, which makes it unsuited for computational purposes if one uses the exhaustive search algorithm.

The selected indicators can be found in Table 3.1 and the detailed description of the indicators can be found in Appendix C.

Selected indicators	Type
Ressurs	Group - manufacturing indicators
SKI.s	Group - manufacturing indicators
Invest	Group - manufacturing indicators
N.0	Group - manufacturing indicators
PrisF.1	Group - manufacturing indicators
svare	Group - employment indicators
sfintj	Group - employment indicators
Ali	Group - other indicators
OTi	Group - other indicators
Lager	Group - other indicators

Table 3.1: Selected indicators.

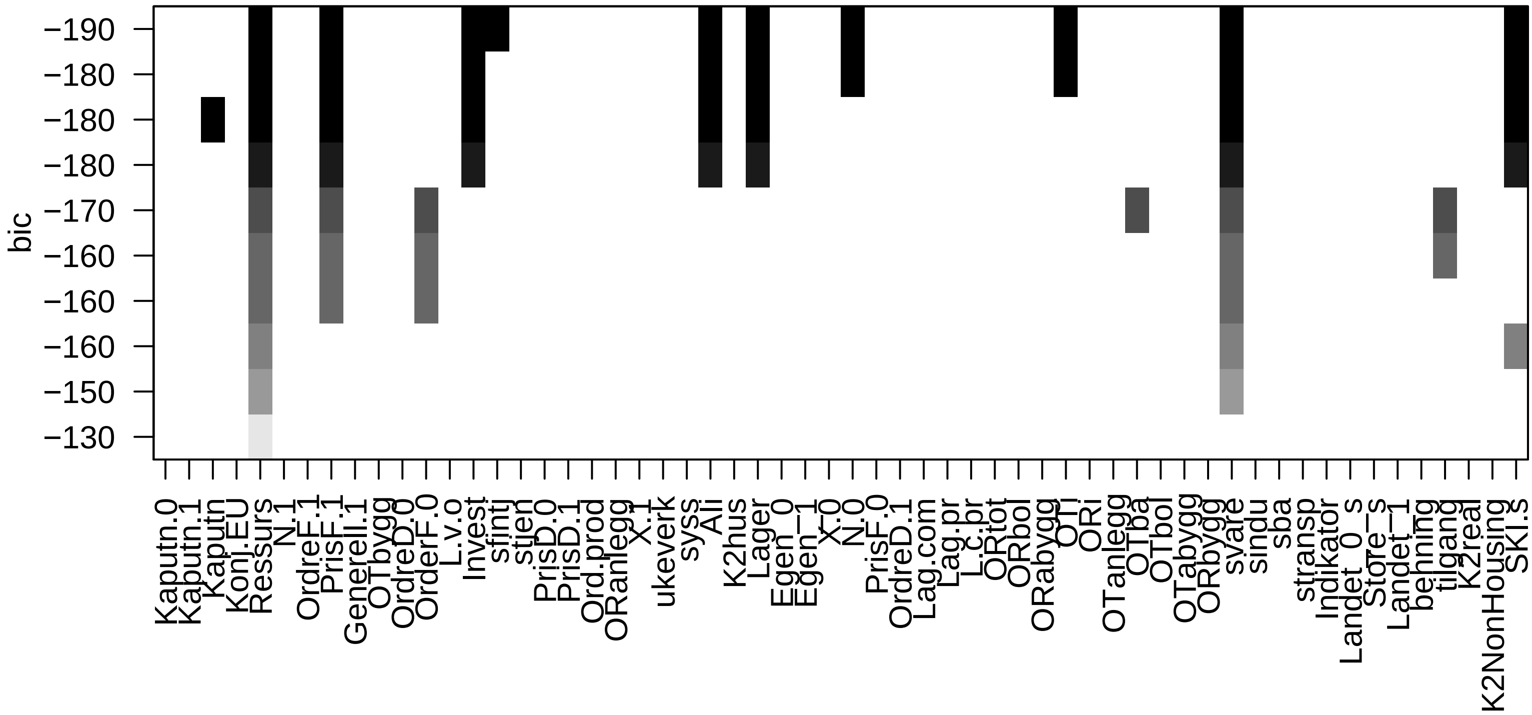

Figure 3.1: Regressors picked by the exhaustive search algorithm and the models corresponding BIC values. The graph displays the stepwise procedure of the exhaustive algorithm and displays the single best model of each size (from 1 to 10 regressors included) when the best model is chosen from all available regressors (57).

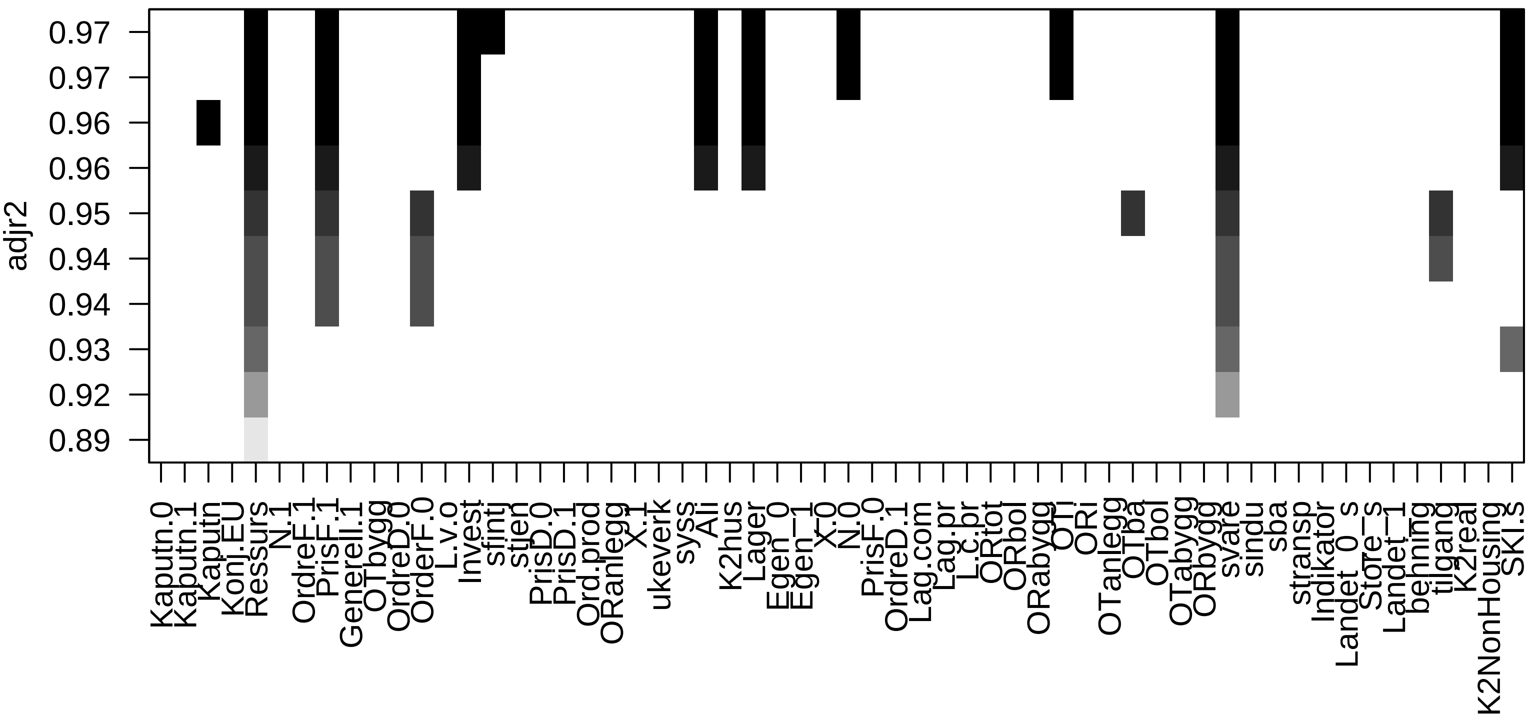

Figure 3.2: Regressors picked by the exhaustive search algorithm and the models corresponding, adjusted R^2 values.

3.2 Forecast performance of the Autoregressive model

This sub chapter describes the results from the AR models which are useful benchmarks for analyzing the performance of the weighting schemes. Figure 3.3 displays the pseudo out-of-sample performance of the best AR model, the specifications of which are described in the figure. Table 3.2 shows the results for lag length $1, ..., 4$, with both recursive and rolling windows (whose principles have been described earlier). The recursive window results are highlighted in *italic* writing and the best performing results for each forecast step are underlined. The forecasts with the lowest RMSE are considered to be the best, if two or more point forecasts for the same step have the same RMSE the one with the lowest bias is selected.

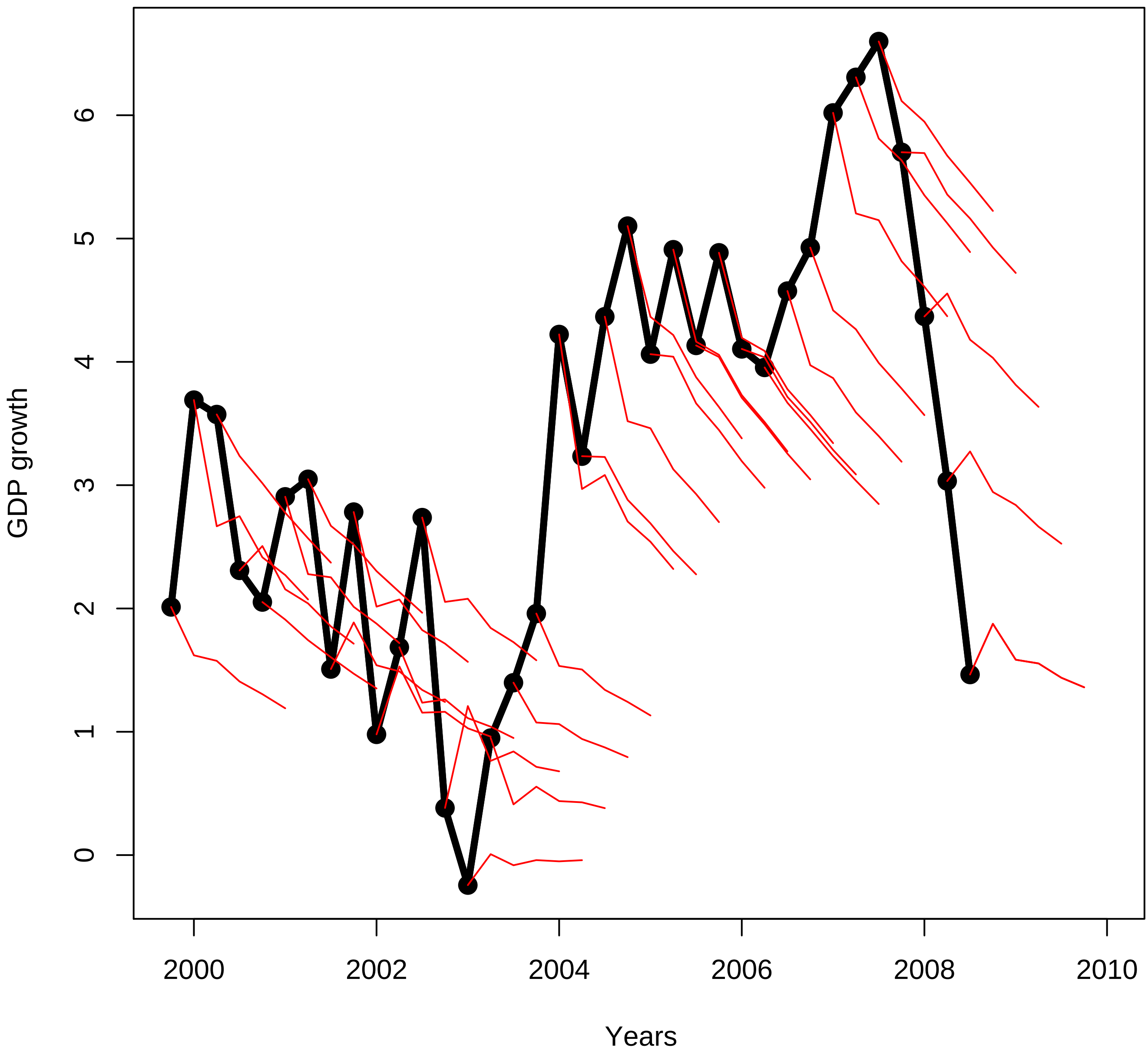

Figure 3.3: Best AR model pseudo out-of-sample performance, recursive window, 2 lags. The thick line shows the actual values of the quarterly GDP growth while the thin line represent the 5 step out-of-sample forecast at each point of time. This is also the structure for the rest of the graphs.

Step	Lag length	Rolling Bias, RMSE	Recursive Bias, RMSE
1	1	0.48 , 1.17	*0.49 , 1.16*
1	2	0.31 , 1.16	*0.30 , <u>1.16</u>*
1	3	0.29 , 1.18	*0.31 , 1.16*
1	4	0.32 , 1.16	*0.32 , 1.16*
2	1	0.89 , 1.64	*0.91 , 1.61*
2	2	0.46 , 1.46	*0.45 , <u>1.45</u>*
2	3	0.42 , 1.49	*0.46 , 1.46*
2	4	0.45 , 1.48	*0.48 , 1.46*
3	1	1.28 , 2.04	*1.29 , 1.99*
3	2	0.71 , 1.76	*0.69 , <u>1.76</u>*
3	3	0.66 , 1.81	*0.71 , 1.76*
3	4	0.70 , 1.82	*0.75 , 1.77*
4	1	1.68 , 2.45	*1.69 , 2.40*
4	2	0.99 , 2.04	*0.96 , <u>2.04</u>*
4	3	0.93 , 2.09	*0.99 , 2.04*
4	4	1.01 , 2.16	*1,05 , 2.08*
5	1	2.07 , 2.74	*2.05 , 2.69*
5	2	1.29 , 2.22	*1.26 , <u>2.21</u>*
5	3	1.25 , 2.24	*1.29 , 2.22*
5	4	1.34 , 2.33	*1.37 , 2.26*

Table 3.2: AR rolling and recursive window result table.

3.3 Forecast performance of Equal weighted combinations

The rest of the chapter displays the results from the different weighting schemes.

Table 3.3, 3.4 and 3.5 show the results for the equal model class weighting schemes for 2VAR, ensembled 2VAR and 3VAR and ensembled 2VAR, 3VAR and 4VAR, while Table 3.6 and 3.7 show equal weighing schemes for ensembled 2VAR and 3VAR and ensembled 2VAR, 3VAR and 4VAR where all models are assigned equal weights independent of which class they belong too. The 2VAR Table is of course the same for both cases.

Step	Lag length	Rolling Bias, RMSE	Recursive Bias, RMSE
1	1	0.27 , 1.03	*0.34 , 1.06*
1	2	0.14 , 1.02	*0.21 , 1.03*
1	3	0.15 , 0.99	*0.20 , 0.99*
1	4	0.17 , 0.99	*0.22 , 0.99*
2	1	0.47 , 1.35	*0.60 , 1.40*
2	2	0.22 , 1.26	*0.28 , 1.27*
2	3	0.21 , 1.21	*0.28 , 1.23*
2	4	0.22 , 1.21	*0.31 , 1.21*
3	1	0.66 , 1.63	*0.84 , 1.70*
3	2	0.35 , 1.54	*0.44 , 1.56*
3	3	0.38 , 1.50	*0.47 , 1.55*
3	4	0.40 , 1.50	*0.53 , 1.53*
4	1	0.89 , 1.89	*1.10 , 1.99*
4	2	0.55 , 1.80	*0.65 , 1.83*
4	3	0.61 , 1.78	*0.69 , 1.85*
4	4	0.64 , 1.81	*0.80 , 1.85*
5	1	1.13 , 2.05	*1.35 , 2.18*
5	2	0.79 , 1.92	*0.89 , 1.97*
5	3	0.88 , 1.92	*0.96 , 2.00*
5	4	0.89 , 1.94	*1.07 , 2.03*

Table 3.3: EW rolling and recursive window result table, 2VAR.

Step	Lag length	Rolling Bias, RMSE	Recursive Bias, RMSE
1	1	0.24 , 0.99	*0.30 , 1.03*
1	2	0.13 , 0.97	*0.18 , 0.99*
1	3	0.12 , 0.97	*0.15 , 0.94*
1	4	0.13 , 0.98	*0.17 , 0.95*
2	1	0.36 , 1.29	*0.51 , 1.35*
2	2	0.17 , 1.18	*0.24 , 1.21*
2	3	0.16 , 1.15	*0.22 , 1.15*
2	4	0.19 , 1.15	*0.24 , 1.14*
3	1	0.52 , 1.60	*0.71 , 1.63*
3	2	0.30 , 1.46	*0.37 , 1.48*
3	3	0.33 , 1.43	*0.39 , 1.47*
3	4	0.32 , 1.42	*0.45 , 1.44*
4	1	0.75 , 1.87	*0.93 , 1.91*
4	2	0.52 , 1.73	*0.57 , 1.74*
4	3	0.54 , 1.73	*0.61 , 1.76*
4	4	0.77 , 1.78	*0.70 , 1.73*
5	1	0.95 , 2.01	*1.15 , 2.07*
5	2	0.71 , 1.86	*0.80 , 1.87*
5	3	0.75 , 1.89	*0.87 , 1.91*
5	4	0.77 , 1.91	*0.96 , 1.91*

Table 3.4: EWC rolling and recursive window result table, ensembled 2VAR and 3VAR.

Step	Lag length	Rolling Bias, RMSE	Recursive Bias, RMSE
1	1	0.23 , 0.97	*0.27 , 1.01*
1	2	0.13 , 0.94	*0.16 , 0.96*
1	3	0.11 , 0.97	*0.11 , <u>0.91</u>*
1	4	0.11 , 0.97	*0.12 , 0.92*
2	1	0.35 , 1.27	*0.45 , 1.33*
2	2	0.17 , 1.15	*0.20 , 1.16*
2	3	0.13 , 1.13	*0.17 , 1.10*
2	4	0.14 , 1.13	*0.18 , <u>1.08</u>*
3	1	0.49 , 1.59	*0.62 , 1.61*
3	2	0.29 , 1.43	*0.32 , 1.43*
3	3	0.29 , 1.43	*0.33 , 1.41*
3	4	0.26 , 1.41	*0.36 , <u>1.38</u>*
4	1	0.68 , 1.85	*0.81 , 1.86*
4	2	0.48 , 1.71	*0.51 , 1.67*
4	3	0.50 , 1.76	*0.55 , 1.70*
4	4	0.48 , 1.79	*0.61 , <u>1.66</u>*
5	1	0.86 , 1.99	*1.01 , 2.00*
5	2	0.66 , 1.83	*0.72 , 1.80*
5	3	0.71 , 1.92	*0.80 , 1.86*
5	4	0.69 , 1.95	*0.86 , <u>1.84</u>*

Table 3.5: EWC rolling and recursive window result table, ensembled 2VAR, 3VAR and 4VAR.

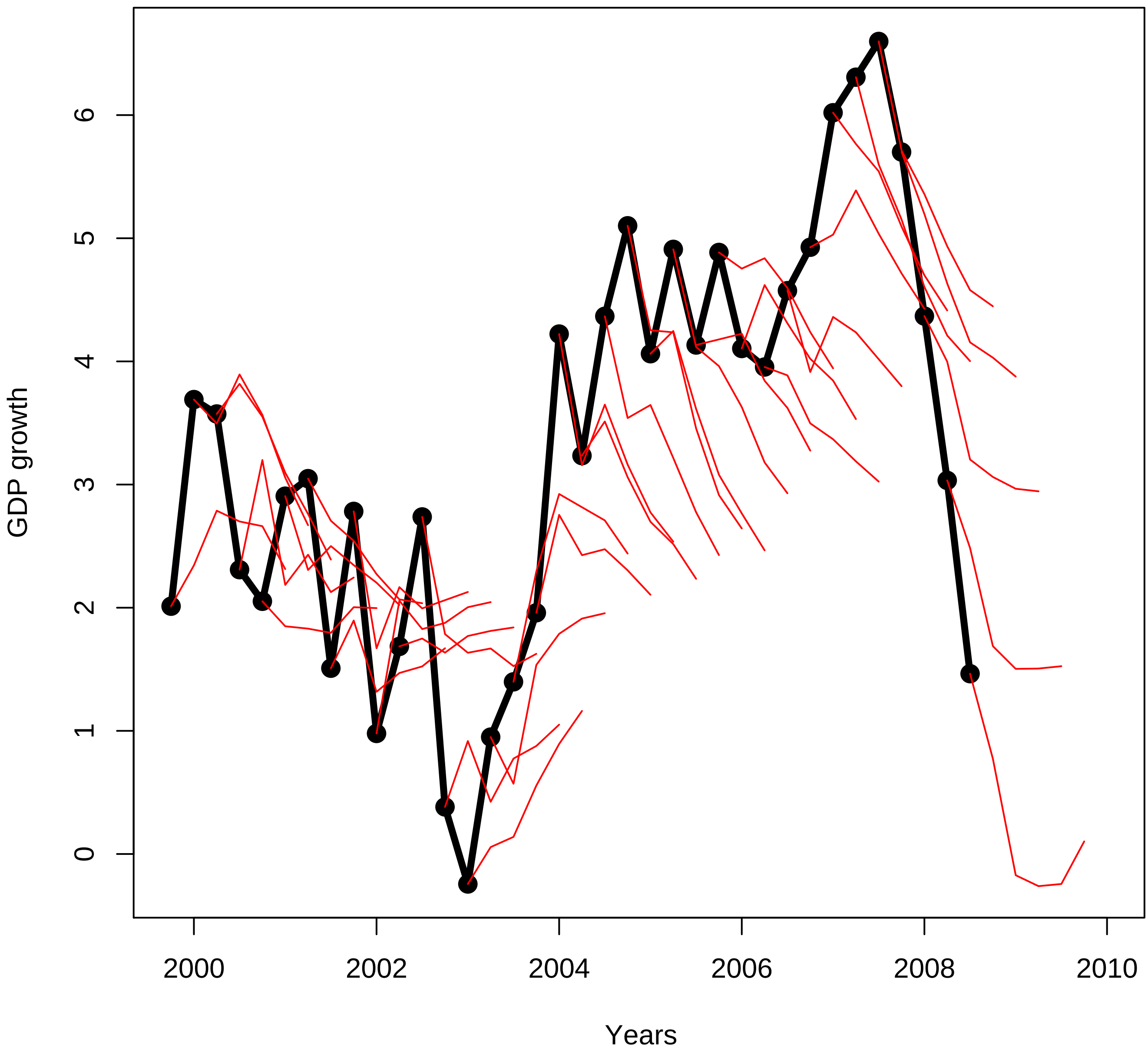

Figure 3.4: Best ensembled EWC model. Ensembled 2VAR, 3VAR and 4VAR models, recursive window, 4lags.

The graph for the best model is picked by evaluating the RMSE values for all forecasting steps. Each step is given the same weight so the best forecast at step 5 counts the same as step 1. The same principle is followed for the other graphs.

Step	Lag length	Rolling Bias, RMSE	Recursive Bias, RMSE
1	1	0.22 , 0.97	*0.27 , 1.01*
1	2	0.12 , 0.95	*0.16 , 0.96*
1	3	0.10 , 0.96	*0.12 , <u>0.92</u>*
1	4	0.11 , 0.97	*0.14 , 0.93*
2	1	0.29 , 1.27	*0.45 , 1.33*
2	2	0.13 , 1.14	*0.21 , 1.17*
2	3	0.12 , 1.12	*0.18 , 1.11*
2	4	0.17 , 1.14	*0.20 , <u>1.10</u>*
3	1	0.44 , 1.60	*0.62 , 1.60*
3	2	0.27 , 1.42	*0.33 , 1.44*
3	3	0.29 , 1.40	*0.34 , 1.42*
3	4	0.27 , <u>1.39</u>	*0.39 , 1.39*
4	1	0.66 , 1.87	*0.82 , 1.85*
4	2	0.50 , 1.70	*0.52 , 1.69*
4	3	0.49 , 1.72	*0.56 , 1.71*
4	4	0.49 , 1.77	*0.64 , <u>1.68</u>*
5	1	0.83 , 2.00	*1.02 , 2.00*
5	2	0.66 , 1.85	*0.73 , 1.82*
5	3	0.67 , 1.90	*0.81 , <u>1.86</u>*
5	4	0.70 , 1.91	*0.89 , 1.86*

Table 3.6: EWM rolling and recursive window result table, ensembled 2VAR and 3VAR.

Step	Lag length	Rolling Bias, RMSE	Recursive Bias, RMSE
1	1	0.22 , 0.94	*0.23 , 0.99*
1	2	0.14 , 0.92	*0.13 , 0.92*
1	3	0.09 , 0.98	*0.07 , <u>0.88</u>*
1	4	0.07 , 0.98	*0.06 , 0.89*
2	1	0.32 , 1.24	*0.37 , 1.31*
2	2	0.16 , 1.11	*0.15 , 1.10*
2	3	0.09 , 1.13	*0.10 , 1.05*
2	4	0.09 , 1.14	*0.09 , <u>1.05</u>*
3	1	0.42 , 1.58	*0.50 , 1.58*
3	2	0.26 , 1.41	*0.26 , 1.37*
3	3	0.23 , 1.45	*0.25 , 1.37*
3	4	0.18 , 1.43	*0.25 , <u>1.34</u>*
4	1	0.58 , 1.84	*0.66 , 1.80*
4	2	0.43 , 1.69	*0.42 , <u>1.60</u>*
4	3	0.45 , 1.81	*0.47 , 1.64*
4	4	0.39 , 1.84	*0.50 , 1.60*
5	1	0.73 , 1.98	*0.82 , 1.93*
5	2	0.60 , 1.80	*0.62 , <u>1.73</u>*
5	3	0.65 , 1.97	*0.71 , 1.82*
5	4	0.59 , 2.05	*0.74 , 1.78*

Table 3.7: EWM rolling and recursive window result table, ensembled 2VAR, 3VAR and 4VAR.

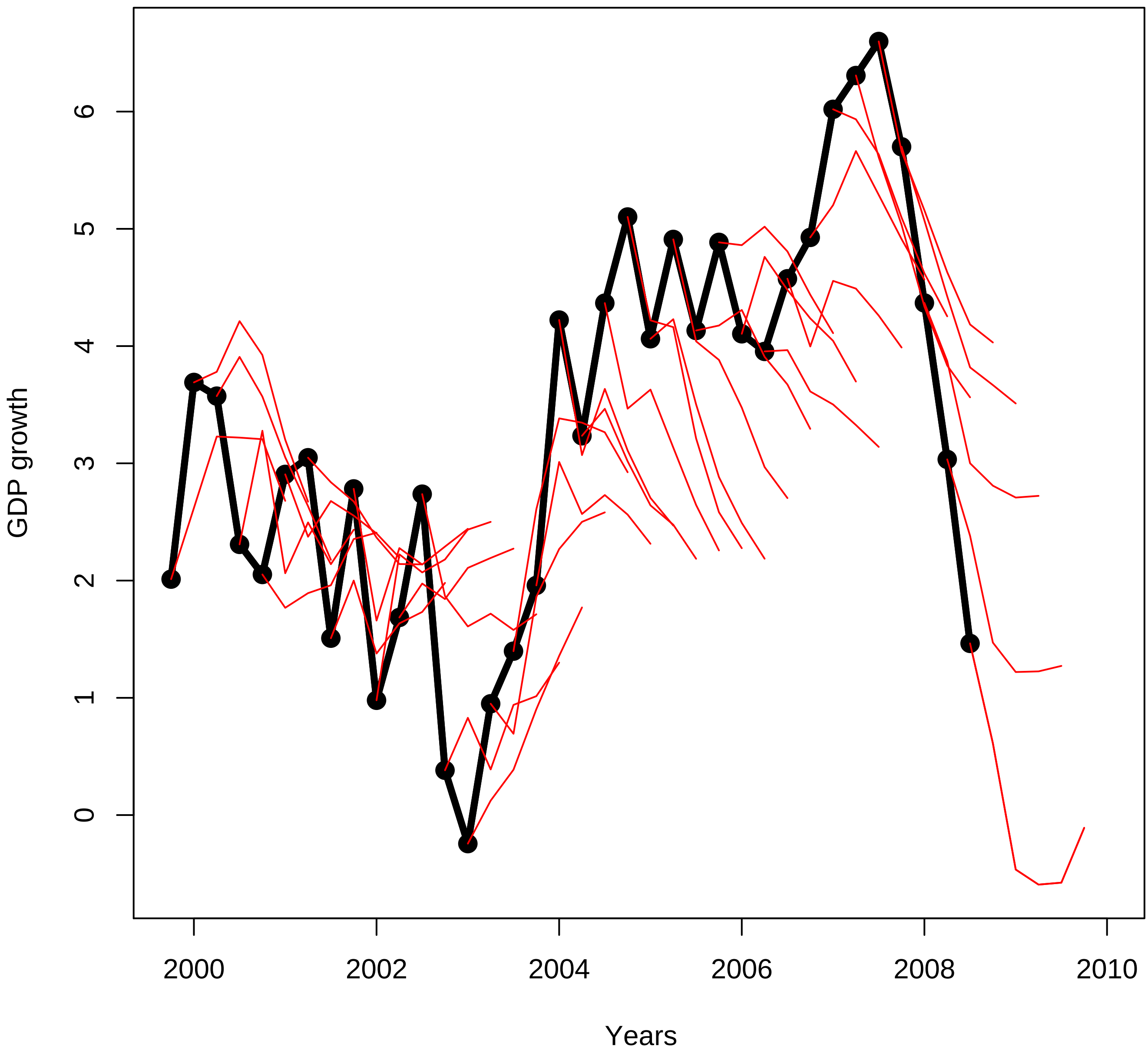

Figure 3.5: Best ensembled EWM model. Ensembled 2VAR, 3VAR and 4VAR models, recursive window, 4lags.

3.4 Forecast performance of Mean squared error weighted combinations

Table 3.8, 3.9 and 3.10 show the results for the mean squared error weighting schemes for 2VAR, ensembled 2VAR and 3VAR and ensembled 2VAR, 3VAR and 4VAR.

Step	Lag length	Rolling Bias, RMSE	Recursive Bias, RMSE
1	1	0.23 , 1.01	*0.30 , 1.04*
1	2	0.09 , 0.98	*0.15 , 1.00*
1	3	0.10 , <u>0.96</u>	*0.14 , 0.96*
1	4	0.13 , 0.96	*0.17 , 0.97*
2	1	0.33 , 1.36	*0.62 , 1.43*
2	2	0.25 , 1.21	*0.31 , 1.27*
2	3	0.28 , 1.14	*0.31 , 1.20*
2	4	0.26 , <u>1.13</u>	*0.36 , 1.17*
3	1	0.54 , 1.75	*0.86 , 1.79*
3	2	0.47 , 1.58	*0.50 , 1.64*
3	3	0.46 , 1.51	*0.51 , 1.62*
3	4	0.47 , <u>1.48</u>	*0.39 , 1.57*
4	1	0.87 , 2.12	*1.17 , 2.17*
4	2	0.79 , 1.97	*0.76 , 1.99*
4	3	0.84 , 1.96	*0.77 , 2.02*
4	4	0.72 , <u>1.94</u>	*0.90 , 2.01*
5	1	1.20 , 2.30	*1.54 , 2.39*
5	2	1.14 , 2.23	*1.12 , <u>2.15</u>*
5	3	1.09 , 2.22	*1.15 , 2.19*
5	4	1.09 , 2.20	*1.29 , 2.22*

Table 3.8: MSEW rolling and recursive window result table, 2VAR.

Step	Lag length	Rolling Bias, RMSE	Recursive Bias, RMSE
1	1	0.20 , 0.97	*0.25* , *1.02*
1	2	0.08 , 0.94	*0.12* , *0.96*
1	3	0.07 , 0.95	*0.10* , <u>*0.93*</u>
1	4	0.11 , 0.95	*0.14* , *0.93*
2	1	0.35 , 1.31	*0.53* , *1.38*
2	2	0.22 , 1.17	*0.27* , *1.20*
2	3	0.22 , *0.26*	1.10 , *1.13*
2	4	0.26 , <u>1.08</u>	*0.32* , *1.11*
3	1	0.52 , 1.69	*0.75* , *1.74*
3	2	0.41 , 1.52	*0.45* , *1.57*
3	3	0.41 , 1.47	*0.48* , *1.55*
3	4	0.48 , <u>1.41</u>	*0.57* , *1.49*
4	1	0.82 , 2.04	*1.04* , *2.09*
4	2	0.70 , <u>1.88</u>	*0.72* , *1.91*
4	3	0.75 , 1.88	*0.77* , *1.93*
4	4	0.68 , 1.85	*0.86* , *1.89*
5	1	1.12 , 2.13	*1.37* , *2.29*
5	2	1.03 , 2.09	*1.08* , <u>*2.08*</u>
5	3	1.04 , 2.13	*1.13* , *2.13*
5	4	1.00 , 2.10	*1.21* , *2.14*

Table 3.9: MSEW rolling and recursive window result table, ensembled 2VAR and 3VAR.

Step	Lag length	Rolling Bias, RMSE	Recursive Bias, RMSE
1	1	0.20 , 0.95	*0.22 , 1.00*
1	2	0.10 , 0.92	*0.11 , 0.93*
1	3	0.06 , 0.95	*0.07 , <u>0.90</u>*
1	4	0.11 , 0.94	*0.10 , 0.91*
2	1	0.36 , 1.29	*0.47 , 1.36*
2	2	0.22 , 1.13	*0.24 , 1.16*
2	3	0.20 , 1.09	*0.22 , 1.08*
2	4	0.24 , <u>1.05</u>	*0.29 , 1.05*
3	1	0.51 , 1.68	*0.68 , 1.71*
3	2	0.39 , 1.48	*0.42 , 1.51*
3	3	0.39 , 1.45	*0.44 , 1.49*
3	4	0.44 , <u>1.36</u>	*0.53 , 1.42*
4	1	0.78 , 2.03	*0.94 , 2.04*
4	2	0.65 , 1.84	*0.69 , 1.83*
4	3	0.71 , 1.86	*0.74 , 1.86*
4	4	0.64 , <u>1.80</u>	*0.81 , 1.80*
5	1	1.08 , 2.22	*1.26 , 2.24*
5	2	0.97 , 2.03	*1.02 , <u>2.01</u>*
5	3	1.00 , 2.11	*1.09 , 2.08*
5	4	0.92 , 2.06	*1.14 , 2.05*

Table 3.10: MSEW rolling and recursive window result table, ensembled 2VAR, 3VAR and 4VAR.

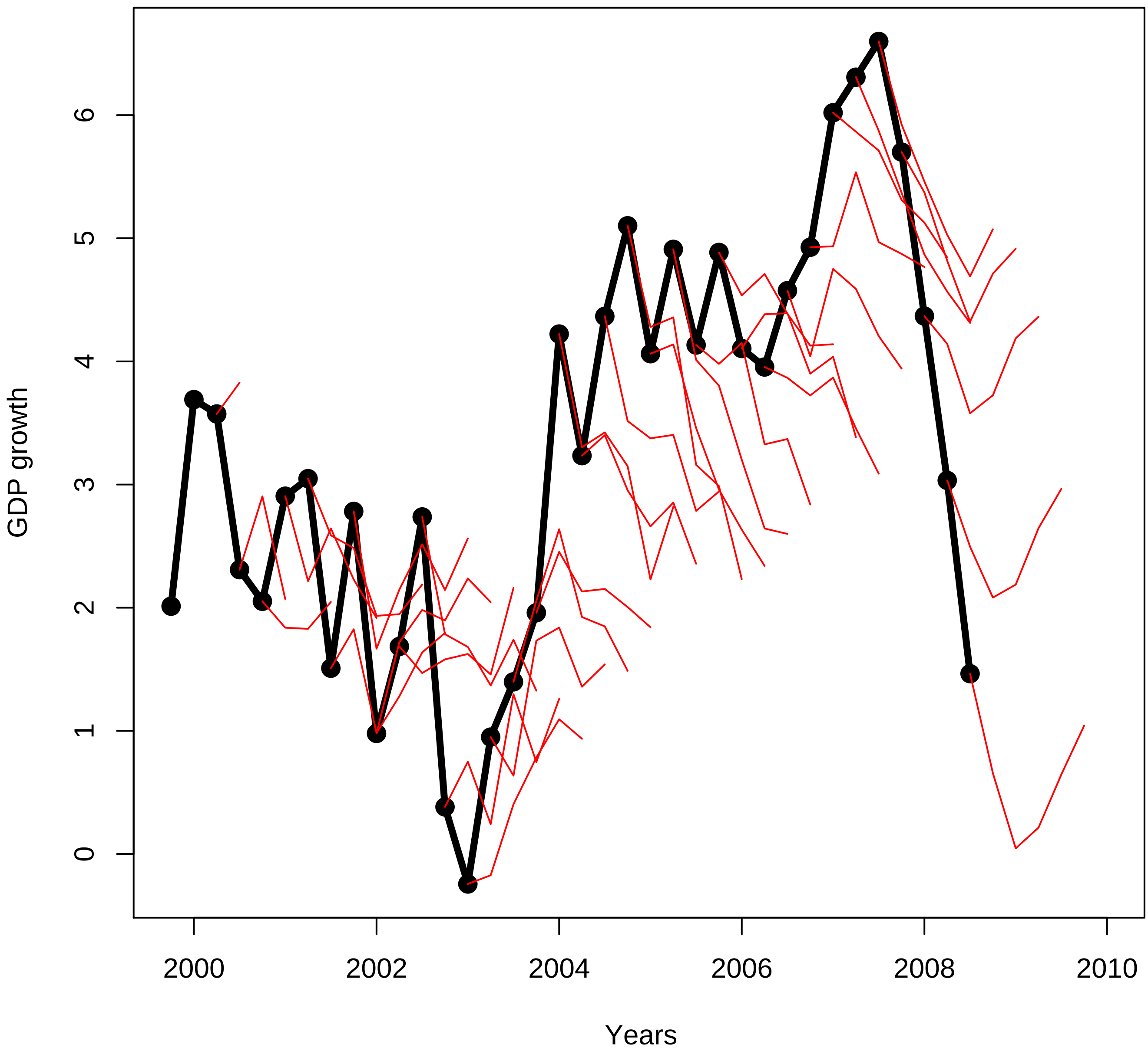

Figure 3.6: Best ensembled MSEW model. Ensembled 2VAR, 3VAR and 4VAR models, rolling window, 4lags.

3.5 Forecast performance of Out-of-sample score-based weighted combinations

Table 3.11, 3.12 and 3.13 show the results for the out-of-sample score-based weighting schemes for 2VAR, ensembled 2VAR and 3VAR and ensembled 2VAR, 3VAR and 4VAR.

Step	Lag length	Rolling Bias, RMSE	Recursive Bias, RMSE
1	1	0.06 , 1.08	*0.09 , 1.08*
1	2	0.02 , 1.07	*0.04 , <u>1.02</u>*
1	3	-0.09 , 1.09	*0.02 , 1.03*
1	4	-0.01 , 1.23	*0.08 , 1.11*
2	1	0.12 , 1.44	*0.17 , 1.41*
2	2	-0.01 , 1.28	*0.07 , <u>1.18</u>*
2	3	-0.18 , 1.32	*0.03 , 1.21*
2	4	-0.11 , 1.44	*0.11 , 1.29*
3	1	0.19 , 1.69	*0.24 , 1.61*
3	2	0.01 , 1.63	*0.14 , <u>1.46</u>*
3	3	-0.25 , 1.66	*0.11 , 1.52*
3	4	-0.19 , 1.77	*0.26 , 1.62*
4	1	0.32 , 1.85	*0.34 , 1.75*
4	2	0.13 , 1.89	*0.29 , <u>1.71</u>*
4	3	-0.10 , 2.01	*0.30 , 1.77*
4	4	-0.06 , 2.12	*0.46 , 1.93*
5	1	0.47 , 1.89	*0.44 , 1.81*
5	2	0.29 , 1.91	*0.44 , <u>1.81</u>*
5	3	0.09 , 2.04	*0.49 , 1.90*
5	4	0.13 , 2.20	*0.64 , 2.03*

Table 3.11: OSSW rolling and recursive window result table, 2VAR.

Step	Lag length	Rolling Bias, RMSE	Recursive Bias, RMSE
1	1	0.13 , 1.13	*0.10 , 1.09*
1	2	0.11 , 1.14	*0.09 , 1.01*
1	3	0.05 , 1.17	*0.13 , <u>0.99</u>*
1	4	0.10 , 1.24	*0.14 , 1.13*
2	1	0.27 , 1.30	*0.17 , 1.42*
2	2	0.21 , 1.24	*0.12 , <u>1.16</u>*
2	3	0.12 , 1.26	*0.17 , 1.20*
2	4	0.16 , 1.32	*0.19 , 1.29*
3	1	0.30 , 1.60	*0.24 , 1.62*
3	2	0.21 , 1.57	*0.20 , <u>1.41</u>*
3	3	0.08 , 1.57	*0.30 , 1.48*
3	4	0.11 , 1.61	*0.34 , 1.69*
4	1	0.45 , 2.07	*0.34 , 1.76*
4	2	0.36 , 2.08	*0.31 , <u>1.68</u>*
4	3	0.24 , 2.13	*0.44 , 1.79*
4	4	0.27 , 2.17	*0.55 , 2.04*
5	1	0.55 , 2.20	*0.44 , 1.83*
5	2	0.46 , 2.21	*0.44 , <u>1.83</u>*
5	3	0.36 , 2.27	*0.59 , 1.96*
5	4	0.38 , 2.33	*0.72 , 2.14*

Table 3.12: OSSW rolling and recursive window result table, ensembled 2VAR and 3VAR.

Step	Lag length	Rolling Bias, RMSE	Recursive Bias, RMSE
1	1	0.14 , 1.09	*0.26 , 1.06*
1	2	0.16 , 1.11	*0.25 , <u>1.02</u>*
1	3	0.12 , 1.11	*0.28 , 1.02*
1	4	0.16 , 1.11	*0.28 , 1.09*
2	1	0.26 , 1.30	*0.33 , 1.32*
2	2	0.22 , 1.22	*0.29 , 1.17*
2	3	0.16 , <u>1.14</u>	*0.33 , 1.20*
2	4	0.25 , 1.19	*0.28 , 1.20*
3	1	0.32 , 1.57	*0.46 , 1.56*
3	2	0.22 , 1.49	*0.43 , <u>1.43</u>*
3	3	0.19 , 1.48	*0.49 , 1.50*
3	4	0.31 , 1.55	*0.52 , 1.57*
4	1	0.48 , 1.97	*0.55 , 1.72*
4	2	0.31 , 1.96	*0.53 , <u>1.66</u>*
4	3	0.38 , 2.06	*0.62 , 1.76*
4	4	0.42 , 2.08	*0.69 , 1.88*
5	1	0.58 , 2.12	*0.61 , 1.78*
5	2	0.40 , 2.10	*0.61 , <u>1.78</u>*
5	3	0.45 , 2.20	*0.71 , 1.90*
5	4	0.52 , 2.25	*0.80 , 2.00*

Table 3.13: OSSW rolling and recursive window result table, ensembled 2VAR, 3VAR and 4VAR.

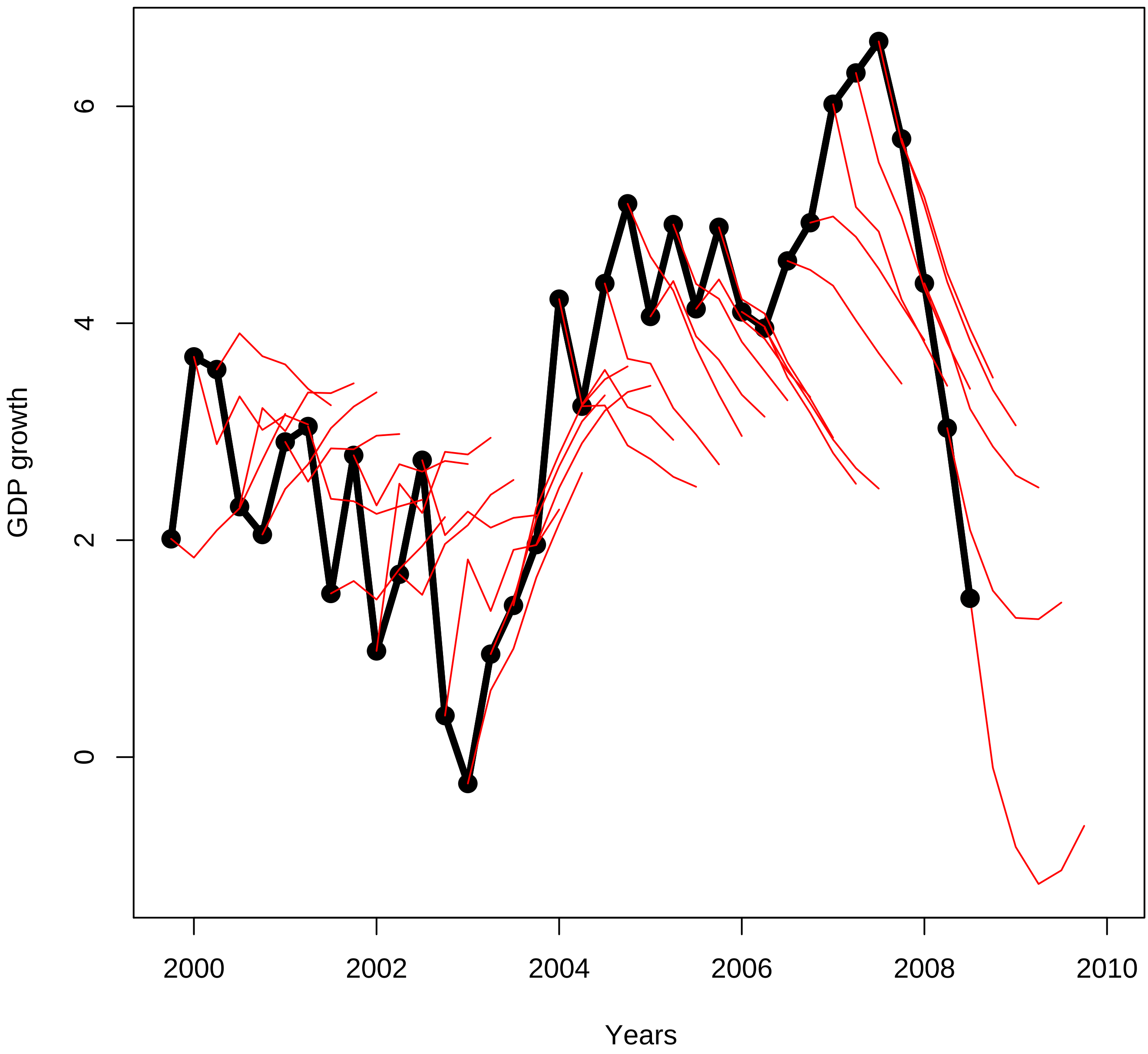

Figure 3.7: Best ensembled OSSW model. Ensembled 2VAR and 3VAR models, recursive window, 2lags.

3.6 Forecast performance of Akaike and Bayesian information criteria weighted combinations

Step	Lag length	Rolling Bias, RMSE	Recursive Bias, RMSE
1	1	0.06 , 1.08	*0.09 , 1.08*
1	2	0.02 , 1.07	*0.04 , <u>1.02</u>*
1	3	-0.09 , 1.09	*0.02 , 1.03*
1	4	-0.01 , 1.23	*0.08 , 1.11*
2	1	0.12 , 1.44	*0.17 , 1.41*
2	2	-0.01 , 1.28	*0.07 , <u>1.18</u>*
2	3	-0.18 , 1.32	*0.03 , 1.21*
2	4	-0.11 , 1.44	*0.11 , 1.29*
3	1	0.19 , 1.69	*0.24 , 1.61*
3	2	0.01 , 1.63	*0.14 , <u>1.46</u>*
3	3	-0.25 , 1.66	*0.11 , 1.52*
3	4	-0.19 , 1.77	*0.26 , 1.52*
4	1	0.32 , 1.85	*0.34 , 1.75*
4	2	0.13 , 1.89	*0.29 , <u>1.71</u>*
4	3	-0.11 , 2.01	*0.30 , 1.77*
4	4	-0.06 , 2.12	*0.46 , 1.93*
5	1	0.47 , 1.89	*0.44 , 1.81*
5	2	0.29 , 1.91	*0.44 , <u>1.81</u>*
5	3	0.08 , 2.04	*0.49 , 1.90*
5	4	0.13 , 2.20	*0.64 , 2.03*

Table 3.14: AIC/BIC rolling and recursive window result table.

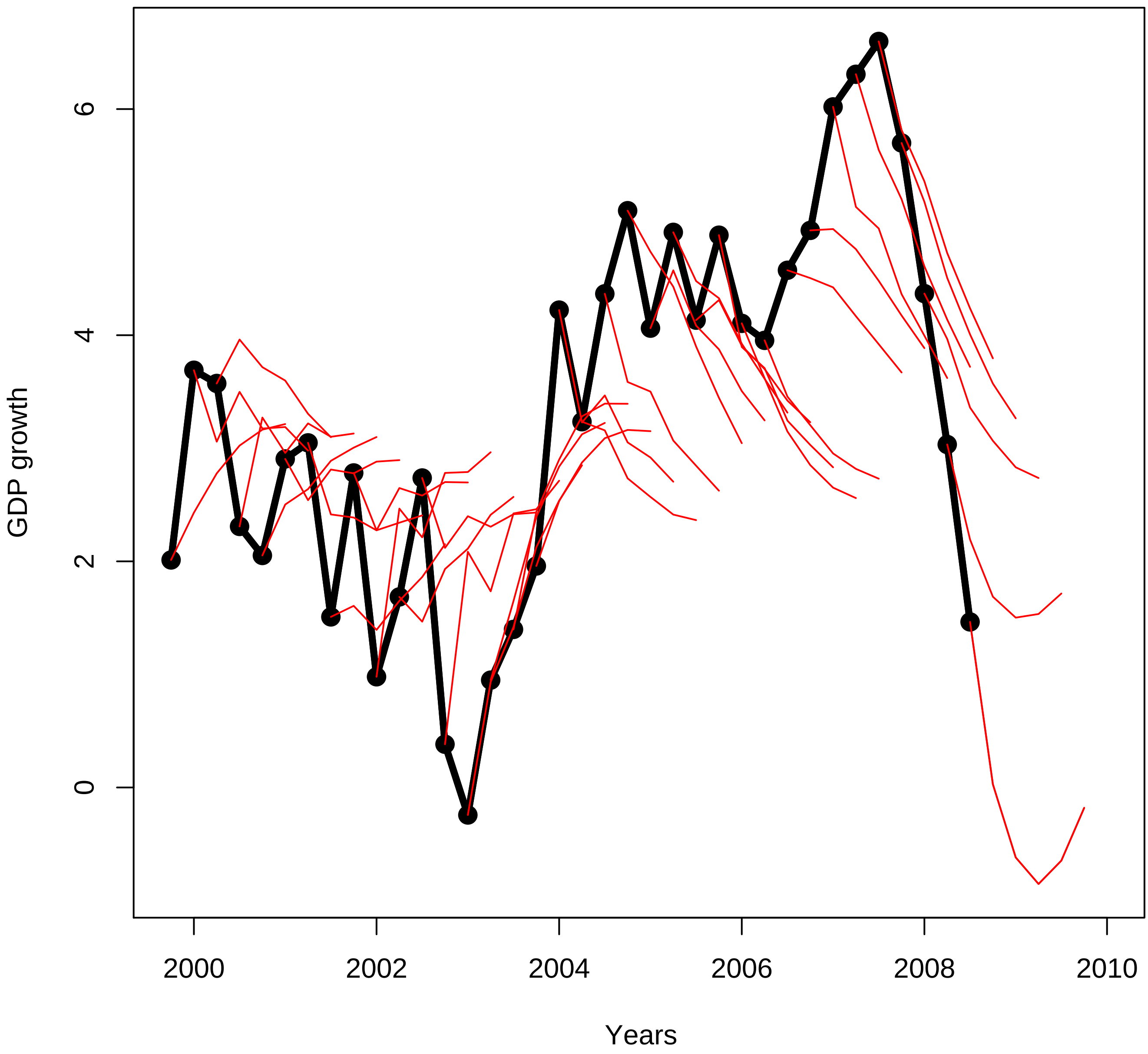

Figure 3.8: Best AICW/BICW model

Because of the negligible differences between AIC an BIC weights (explained in A.2). One gets the same graph for 2VAR, ensembled 2VAR and 3VARs and ensembled 2VAR, 3VAR and 4VARs for AIC and BIC weights.

3.7 Discussion of results

One of the first things that can be noticed from the result tables is that all of the forecast combinations (both the ensembled and non ensembled weighting schemes) for EWM, EWC, MSEW, AICW and BICW perform better than a simple AR model with 1 to 4 lags. On the other hand, none of the weighted averages are even close to beating the performance of the individual *ex post* best models listed in Appendix D which do not deal with model uncertainty *ex ante*.

3.7.1 Out-of-sample performance

When it comes to out-of-sample performance the EWC and MSEW perform quite similarly. The most important result is that ensembling models together (making use of 3VARs and 4VARs) improves the forecast performance for all steps ahead (1 to 5). This means that for both EWC and MSEW 2VAR and 3VAR models ensembled together perform better than only 2VAR models weighted together and 2VAR, 3VAR and 4VAR models weighted together outperform the smaller model ensembles for all forecast steps. This is an important result since all the model classes had equal weights in the ensembled weighting schemes prior to the out-of-sample performance test. This shows that there is a gain in performance from exploring the three and four variable VAR framework and that the GDP forecasts relying on two variable VAR models have limited predictive power compared to the alternative. Output prediction based on lags of GDP and only one single survey indicator is therefore limited and this result emphasizes the assumption that if one chooses to forecast the near-term behavior of the economy (output prediction) with non-structural reduced form models it is advantageous to use more than one survey indicator.

The pattern of the out-of-sample performance for the OSS weighted averages is a little different than for the EWC and MSEW averages. Although one can see clear improvement from only using two variable VAR models to using ensembled two and three variable VARs, the improvement in also including four variable VAR models in the ensembled average is not so clear (see Tables 3.12 and 3.13), since there is improvement for step 2,3 and 5, but not step 1 and 3. When the OSS weighted averages are compared to the other out-of-sample estimated weighted averages that are ensembled regarding to model class, their performance for step 4 and 5 is better than for EWC and MSEW weighted averages and the step 3 performance is very similar. Out-of-sample performance of the first and second step is however poorer for OSSW than for EWC and MSEW averaged point forecasts.

When it comes to the out-of-sample forecast performance of the averages with in-sample derived weights, we found that the differences between weights derived from in-sample BIC and AIC was negligible and hence their out-of-sample forecast performances were comparable. These schemes attached very little weight to the larger (3VAR and 4VAR) models. These averages perform worse with respect to the quadratic RMSE loss function than the other ensembles mentioned above (see Table 3.14). This is because the BIC and AIC weight when given equal prior probabilities, attach very little weight to the larger models although the other weighting schemes have shown that adding the larger models to the ensemble improves the out-of-sample performance.

The simple mean combination forecasts, where all models are given equal weight independent of model class (EWM), have the overall best out-of-sample performance, where the grand ensemble of 2VAR, 3VAR and 4VAR models outperforms the more sophisticated weighting schemes for all forecast steps ($h = 1, ..., 5$) and confirms the results of Stock and Watson (2004). Nevertheless its interesting to note that the BIC/AIC weighted 2VAR ensemble outperforms the equal weighted 2VAR ensemble out-of-sample.

Bias: Clear patterns also emerge with respect to how the different combined forecasts are biased. For equal weighted models and mean squared error weighted, both for ensembled and non ensembled averages, the pattern is very similar. All averaged point forecasts are positively biased with no exception. This means that all weighted averages of this form based on the 10 selected survey indicators, underestimate quarterly GDP growth in the pseudo out-of-sample framework. The weighted forecasts (EWC, EWM and MSEW) also have similar bias patterns for the different lag lengths. Forecasts made with lag length 2 and 3 are the least biased for weighted 2VARs, ensembled 2VAR and 3VARs, and ensembled 2VAR, 3VAR and 4VARs. While forecasts made with lag length 1 have the highest bias for all weighting schemes, these results also hold for the AR model (see Figure; 3.2, 3.3, 3.4, 3.5, 3.8, 3.9, 3.10).

The bias pattern on the BIC, AIC, and OSS weighted schemes are a little different than for EW (C and M) and MSEW who have very similar bias patterns. The most important observation is that when one compares the different ensembles, one finds that Bayesian, Akaike and out-of-sample score-based weighted point forecasts are less biased than the other weighting schemes. Another interesting observation is that some of the AIC, BIC and OSS weighted point forecasts of the 2VAR models are negatively biased, which indicates that they overestimate GDP growth out-of-sample, this has not been observed for any of the other weighting schemes and model ensembles.

Chapter 4

Conclusion and direction for future research

In the introduction of this thesis the following questions were proposed: Is it limiting to focus on bivariate vector autoregressive models when forecasting GDP growth? Can there be any gain in exploring the three and four variable VAR framework to forecast GDP growth? What combinations of survey indicators perform well in forecasting GDP growth at short horizons? How can one efficiently weight individual forecasting models to deal with model uncertainty?

The results in this thesis clearly indicate that it is limiting to focus only on bivariate vector autoregressive models when one uses survey indicators to forecast GDP growth and, there is gain in exploring the three and four variable VAR framework. From the experience gained by performing this exercise one can say that 4VAR models implicitly had the highest posterior weight. Better performing models for short term evaluation of GDP growth than the bivariate have therefore been proposed in this thesis. The combinations of indicators that perform well have mostly been manufacturing indicators (see Table C.1) if one makes judgment from the the *ex post* best models (see Appendix D). From the 10 indicators that were selected by RegSubSets (they can be found in Table 3.1) and were used in the weighting exercise, half were manufacturing indicators. These indicators were selected from the full dataset which contains 57 indicators that are described in Appendix C. One of the indicators (SKI.s - a compound industry indicator) was a part of the 10 selected as well as one of the most frequent regressors in the *ex post* best models, but to create good forecast models according to the RMSE criteria it had to be combined with other survey indicators as well. The indicators “Ressurs” - an indicator for industrial resource scarcity, and “svare” - which indicates change in employment for the construction sector and domestic factories, were the two regressors that the exhaustive search algorithm picked most frequently. The algorithm searched for the best possible model (using GDP growth and the survey indicators) adding one indicator at the time. When the best model with one indicator was found, then a search over the models that add one more regressor was conducted, and the best model with two indicators was recovered. This continued until the best model with GDP growth and 10 indicators was found. The indicator "Ressurs" was a part of all the models that were recovered by the algorithm, hence it was a part of the model containing GDP growth and only one indicator as well as in all the other 9 models that were returned. The indicator "svare" was part of all the models except the one that contained only two variables.

Neither the maximum likelihood based weighted averages nor the mean squared error weighted averages outperformed the simple mean combination of forecasts. One can therefore conclude that in the case of this project, the most efficient way to weight individual forecasting models, to deal with model uncertainty has been a simple mean combination of forecasts.

Comparing the performance of Bayesian/Akaike and out-of-sample score based weights with equal and mean squared error weights has shown that there are good reason to continue the research on maximum likelihood based weights when weighting together forecasts made from combinations of survey indicators. Although the AIC, BIC and OSS weighted point forecasts have a poorer out-of-sample performance on first and second forecast step, their performance is better than MSEW weighting schemes for the 4th and 5th forecast step and also give less biased forecast averages.

For future research based on the results of this project it could be interesting to perform the same exercise by using the direct forecasting method instead of the iterated using the "local linear projection" method advocated by Jorda (2005). Each of the two methods has its advantages and disadvantages. If the VAR model provides a good approximation to the correlations in the data, the iterated forecast method will tend to produce more precise forecasts, in this case create models with lower out-of-sample RMSE values, because the iterated forecast method uses coefficient estimators form the estimators of previous periods. If on the other hand the VAR is incorrectly specified, extrapolating these forecasts by iterating may compound biases (and biased forecast averages are really the case when one looks at the results in this project). Hence if the VAR models are poor, then the direct forecasting method may be more accurate.

Another interesting subject for future research could be more sophisticated specification of prior model and class probabilities in the different weighting and ensembling schemes. Since we have learned from this project that adding three and four variable model classes to two variable classes improves model averaged point forecast performance. One could use some of the priors suggested by Fernandez et al. (2001, page. 15) on the individual models to further improve the performance of the maximum likelihood based weighting schemes.

Bibliography

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: International Symposium on Information Theory, 2 nd, Tsahkadsor, Armenian SSR. pp. 267–281.

Akaike, H., 1974. A new look at the statistical model identification. Automatic Control, IEEE Transactions on 19 (6), 716–723.

Akaike, H., 1978. On the likelihood of a time series model. The Statistician 27, 217–235.

Akaike, H., 1979. A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. Biometrika 66 (6), 237–242.

Akaike, H., 1983. Information measures and model selection. Proceedings of the 44th Session of the International Statistical Institute 1, 277–291.

Banerjee, A., Marcellino, M., 2006. Are there any reliable leading indicators for US inflation and GDP growth? International Journal of Forecasting 22, 137–151.

Banerjee, A., Marcellino, M., Masten, I., May 2003. Leading indicators for euro area inflation and GDP growth. CEPR Discussion Papers 3893, C.E.P.R. Discussion Papers.

Bates, J. M., Granger, C. W. J., Dec 1969. The combination of forecasts. Operational Research Society Vol. 20 (No. 4), pp. 451–468.

Burnham, K. P., Anderson, D. R., 2002. Model Selection and Multimodel Inference, 2nd Edition. Springer.

Chipman, H., George, E., McCulloch, R., 2000. Hierarchical priors for Bayesian cart shrinkage. Statistics and Computing (10), 17–24.

Dovern, J., Jul. 2006. Predicting GDP components. Do leading indicators increase predictability? Kiel Advanced Studies Working Papers 436, Kiel Institute for the World Economy.

Dresse, L., Nieuwenhuyze, C. V., Mar. 2008. Do survey indicators let us see the business cycle? A frequency decomposition. Research series 200803-31, National Bank of Belgium.

Eklund, J., Karlsson, S., 2007. Forecast combination and model averaging using predictive measures. Econometric Reviews 26 (2-4), 329–363.

Elliott, G., Granger, C., Timmermann, A. (Eds.), 2006. Handbook of Economic Forecasting, 1st Edition. Vol. 1 of Handbook of Economic Forecasting. Elsevier.

Fernandez, C., Ley, E., Steel, M. F. J., February 2001. Benchmark priors for Bayesian model averaging. Journal of Econometrics 100 (2), 381–427.

Granger, C. W. J., Ramanathan, R., 1984. Improved methods of combining forecasts. Journal of Forecasting (3), 197–204.

Hamilton, J. D., 1994. Time Series Analysis, 1st Edition. Princeton University Press.

Hoeting, J. A., Madigan, D., Raftery, A. E., Volinsky, C. T., 1999. Bayesian model averaging: A tutorial. Statistical Science 14 (4), 382–401.

Jorda, O., March 2005. Estimation and inference of impulse responses by local projections. American Economic Review 95 (1), 161–182.

Kascha, C., Ravazzolo, F., Dec. 2008. Combining inflation density forecasts. Working Paper 2008/22, Norges Bank.

Lucas, R. J., January 1976. Econometric policy evaluation: A critique. Carnegie-Rochester Conference Series on Public Policy 1 (1), 19–46.

Madigan, D., Raftery, A. E., December 1994. Model selection and accounting for model uncertainty in graphical models using Occams's window. Journal of the American Statistical Association 89 (428), 1535–1546.

Marcellino, M., 2008. A linear benchmark for forecasting GDP growth and inflation? Journal of Forecasting 27 (4), 305–340.

Marcellino, M., Stock, J. H., Watson, M. W., 2006. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. Journal of Econometrics 135 (1-2), 499–526.

Miller, A. J., 2002. Subset Selection in Regression, 2nd Edition. Chapman and Hall.

Min, C.-K., Zellner, A., March 1993. Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates. Journal of Econometrics 56 (1-2), 89–118.

Pfaff, B., 2008. VAR, SVAR and SVEC models: Implementation of vars within R package. Journal of Statistical Software 27 (4).

Rissanen, J., 1986. Stochastic complexity and modeling. The Annals of Statistics (14), 1080–1100.

Schwarz, G., 1978. Estimating the dimension of a model. The Annals of Statistics 2 (6), 461–464.

Sims, C. A., January 1980. Macroeconomics and reality. Econometrica 48 (1), 1–48.

SSB, 2008. Konjunkturbarometeret. `http://www.ssb.no/emner/08/05/10/kbar`.

Stock, J. H., Watson, M. W., 2004. Combination forecasts of output growth in a seven-country data set. Journal of Forecasting 23 (6), 405–430.

Wei, C. Z., 1992. On predictive least squares principles. The Annals of Statistics (20), 1–42.

Wright, J. H., October 2008. Bayesian model averaging and exchange rate forecasts. Journal of Econometrics 146 (2), 329–341.

Wright, J. H., 2009. Forecasting US inflation by Bayesian model averaging. Journal of Forecasting 28 (2), 131–144.

Appendix A

Mathematical supplement

A.1 Similarities between AIC, BIC and OSS weights

This section shows why AIC weights and BIC weights give the same results when and why both equal log-likelihood weights (OSSW) for fixed-sized models when the models are first weighted together with models of the same size.

$$AIC_i = -2(\Lambda_i) + 2K_i$$

$$BIC_i = -2(\Lambda_i) + K_i log(T_i - p_i)$$

while

$$AIC_{min} = -2(\Lambda_{min}) + 2K_{min}$$

$$BIC_{min} = -2(\Lambda_{min}) + K_{min} log(T_{min} - p_{min})$$

and since $K_i = K_{min}$, $p_i = p_{min}$ and $T_k = T_{min}$ in my model classes both $AIC_i - AIC_{min}$ and $BIC_i - BIC_{min}$ become $2(\Lambda_{min} - \Lambda_i)$. Since the differences between AIC and BIC weights only stem from the size of the penalty for having a large K_i. Next we show why both AIC and BIC weights are equal to weights derived from the log-likelihood for models of fixed parameter size:

$$\omega_{t+h|t}^{(i)} = \frac{exp\left\{-\frac{1}{2}\Delta_{t+h|t}^{(i)}(BIC)\right\}}{\sum_{j=1}^{N} exp\left\{-\frac{1}{2}\Delta_{t+h|t}^{(j)}(BIC)\right\}} = \frac{exp\left\{-\frac{1}{2}\Delta_{t+h|t}^{(i)}(AIC)\right\}}{\sum_{j=1}^{N} exp\left\{-\frac{1}{2}\Delta_{t+h|t}^{(j)}(AIC)\right\}} =$$

$$\frac{exp\left\{-\frac{1}{2}\left(AIC_{t+h|t}^{(i)} - min(AIC)_{t+h|t}\right)\right\}}{\sum_{j=1}^{N} exp\left\{-\frac{1}{2}\left(AIC_{t+h|t}^{(j)} - min(AIC)_{t+h|t}\right)\right\}} =$$

$$\frac{exp\left\{-\frac{1}{2}\left(AIC_{t+h|t}^{(i)}\right)\right\} exp\left\{\frac{1}{2}\left(min(AIC)_{t+h|t}\right)\right\}}{\sum_{j=1}^{N} exp\left\{-\frac{1}{2}\left(AIC_{t+h|t}^{(j)}\right)\right\} exp\left\{\frac{1}{2}\left(min(AIC)_{t+h|t}\right)\right\}} =$$

$$\frac{exp\left\{-\frac{1}{2}\left(-2\Lambda^{(i)}_{t+h|t}-2K_i\right)\right\}}{\sum_{j=1}^{N} exp\left\{-\frac{1}{2}\left(-2\Lambda^{(j)}_{t+h|t}-2K_j\right)\right\}}=\frac{exp\left\{\Lambda^{(i)}_{t+h|t}\right\}}{\sum_{j=1}^{N} exp\left\{\Lambda^{(j)}_{t+h|t})\right\}}. \tag{A.1}$$

Because $K^{(i)}_{t+h|t}=K^i=K^j$ when the weighted models are of a fixed size. We can therefore write:

$$\frac{exp\left\{-\frac{1}{2}\Delta^{(i)}_{t+h|t}(BIC)\right\}}{\sum_{j=1}^{N} exp\left\{-\frac{1}{2}\Delta^{(j)}_{t+h|t}(BIC)\right\}}=\frac{exp\left\{-\frac{1}{2}\Delta^{(i)}_{t+h|t}(AIC)\right\}}{\sum_{j=1}^{N} exp\left\{-\frac{1}{2}\Delta^{(j)}_{t+h|t}(AIC)\right\}}=\frac{exp\left\{\Lambda^{(i)}_{t+h|t}\right\}}{\sum_{j=1}^{N} exp\left\{\Lambda^{(j)}_{t+h|t}\right\}}. \tag{A.2}$$

A.2 Negligible differences between AIC and BIC weights

Lets call these in-sample weights $\omega^{(i)}_{AIC}$ and $\omega^{(i)}_{BIC}$. If we ensemble together 2 and 3VAR models, the weights for an individual 2VAR model i can be written like this:

$$\omega^{(i)}_{AIC}=\frac{exp\left\{-\frac{1}{2}AIC_i\right\}}{\sum_{j}^{N_2} exp\left\{-\frac{1}{2}AIC_j\right\}+\sum_{j}^{N_3} exp\left\{-\frac{1}{2}AIC_j\right\}}=$$

$$\frac{exp\left(\Lambda_i\right)}{exp\left(K_2\right)\left(\frac{1}{exp(K_2)}\sum_{j}^{N_2} exp\left(\Lambda_j\right)+\frac{1}{exp(K_3)}\sum_{j}^{N_3} exp\left(\Lambda_j\right)\right)}=$$

$$\frac{exp\left(\Lambda_i\right)}{\sum_{j}^{N_2} exp\left(\Lambda_j\right)+\frac{exp(K_2)}{exp(K_3)}\sum_{j}^{N_3} exp\left(\Lambda_j\right)}=$$

$$\frac{exp\left(\Lambda_i\right)}{\sum_{j}^{N_2} exp\left(\Lambda_j\right)+exp\left(K_2-K_3\right)\sum_{j}^{N_3} exp\left(\Lambda_j\right)}. \tag{A.3}$$

To calculate $\omega^{(i)}_{BIC}$ we get:

$$\omega^{(i)}_{BIC}=\frac{exp\left\{-\frac{1}{2}BIC_i\right\}}{\sum_{j}^{N_2} exp\left\{-\frac{1}{2}BIC_j\right\}+\sum_{j}^{N_3} exp\left\{-\frac{1}{2}BIC_j\right\}}=$$

$$\frac{exp\left(\Lambda_i\right)}{exp\left(\frac{K_2}{2}log(T_2-p_2)\right)\left(\frac{1}{exp\left(\frac{K_2}{2}log(T_2-p_2)\right)}\sum_{j}^{N_2} exp\left(\Lambda_j\right)+\frac{1}{exp\left(\frac{K_3}{2}log(T_3-p_3)\right)}\sum_{j}^{N_3} exp\left(\Lambda_j\right)\right)}$$

and since $T_2=T_3=T$, $p_2=p_3=p$ the equation becomes

$$\omega^{(i)}_{BIC}=\frac{exp\left(\Lambda_i\right)}{\sum_{j}^{N_2} exp\left(\Lambda_j\right)+exp\left(\frac{log(T-p)}{2}\left(K_2-K_3\right)\right)\sum_{j}^{N_3} exp\left(\Lambda_j\right)}. \tag{A.4}$$

The only difference between the weight for the 2VAR model i is that the AIC weight has (K_2-K_3) in the denominator, while the BIC weight has $\left(\frac{log(T-p)}{2}\left(K_2-K_3\right)\right)$. In our framework (for which K still is the number of parameters estimated in the system) and

$K_2 = 22$, $K_3 = 48$, $T = 63$ and p is the lag length of the VAR. If we calculate these different values with lag length equaling 4 we get:

$$exp\,(K_2 - K_3) \approx 5.11 \times 10^{-12}$$

$$exp\left(\frac{log\,(T-p)}{2}\,(K_2 - K_3)\right) \approx 9.53 \times 10^{-24}.$$

These numbers are so small that they make the denominator almost equal for both cases. If we want to compare the weights for an individual 3VAR model i we can by the same calculation method obtain:

$$\omega_{BIC}^{(i)} = \frac{exp\,(\Lambda_i)}{exp\left(\frac{log(T-p)}{2}\,(K_3 - K_2)\right)\sum_j^{N_2} exp\,(\Lambda_j) + \sum_j^{N_3} exp\,(\Lambda_j)}. \tag{A.5}$$

$$\omega_{AIC}^{(i)} = \frac{exp\,(\Lambda_i)}{exp\,(K_3 - K_2)\sum_j^{N_2} exp\,(\Lambda_j) + \sum_j^{N_3} exp\,(\Lambda_j)}. \tag{A.6}$$

and

$$exp\,(K_3 - K_2) \approx 1.96 \times 10^{11}$$

$$exp\left(\frac{log\,(T-p)}{2}\,(K_3 - K_2)\right) \approx 1.05 \times 10^{23}.$$

Again we can see that the part of the denominator that is related to the 2VAR models becomes almost infinitely larger than the one that belongs to 3VARs and the difference AIC and BIC weights becomes negligible.

To test if one could detect a difference in the weighted forecasts when using the AIC and BIC method $\frac{1}{2}log\,(T-p) = 2$ was used in the program code. Then one could detect differences in the fifth decimal so the bias and RMSE differed by 0.00005. One can also notice that both AIC and BIC weight punish the larger models very hard when the models are given equal prior probability and that almost all the weight is placed on the 2VAR models while the results from 3VAR models get no weight in the ensemble at all. The same conclusions can be drawn when 2VAR, 3VAR and 4VAR models are ensembled together and forecasts are weighted with AIC and BIC weights. When one sets the lag length of the VAR to its minimum (1) and gets $K_3 = 21$ and $K_2 = 10$ one still can not detect any difference in the third decimal.

Appendix B

Model space

All categories of VAR models were run with lag length 1 to n, where n is set to 4 after testing higher values and discovering that models with $n > 4$ did not give better out-of-sample forecasts and caused problems for the computation of Akaike information criteria (AIC) and Bayesian information criteria (BIC) weights[1]. The models were also estimated with both rolling and recursive windows.

[1]OSS weights is short for Out-of-sample score-based weights

VAR size	Model name	Individual model weights
1	AR	None
2	2VAR	None
3	3VAR	None
4	4VAR	None
2	2VAR	OSS
2	2VAR	Equal
2	2VAR	MSE
2	2VAR	BIC
2	2VAR	AIC
3	Ensembled 2 and 3VAR	OSS
3	Ensembled 2 and 3VAR	Equal
3	Ensembled 2 and 3VAR	MSE
3	Ensembled 2 and 3VAR	BIC
3	Ensembled 2 and 3VAR	AIC
4	Ensembled 2, 3 and 4VAR	OSS
4	Ensembled 2, 3 and 4VAR	Equal
4	Ensembled 2, 3 and 4VAR	MSE
4	Ensembled 2, 3 and 4VAR	BIC
4	Ensembled 2, 3 and 4VAR	AIC

Table B.1: List of models

Appendix C

Dataset description

This part of the appendix offers a brief explanation of the indicators used together with GDP growth in the forecasting process. The tables contain information about the time series such as name, explanation of the series, and transformations. The letters in the table have the following explanations;

- q means quarterly
- m means monthly
- t means trended series
- s means seasonally adjusted
- n means not seasonally adjusted
- diff means diffusion index
- w means weighted average
- agg means aggregate
- pres is present
- prev is previous
- break is break in time series from 2006

The diffusion index is calculated in the following way: (share of those who answered the survey: bigger) + 0.5×(share of those who answered the survey: unchanged). The diffusion index has a turning point around 50. A score of 50 means that the number of participants expecting positive growth in the variable equals the number of participants expecting negative growth.

There are in total 57 potential indicators in the dataset: 25 manufacturing, 6 trend, 10 construction and factory order reserves and supply indicators, 10 employment indicators, and 6 others.

Series name	Explanation	Transformations
N.0	Average occupation, change pres vs. prev q	diff, t
N.1	Average occupation, change next vs. pres q	diff, t
X.0	Total industrial production volume, change pres vs. prev q	diff, t
X.1	Total industrial production volume, change next vs. pres q	diff, t
Kaputn	Capacity utilization level given present production, q	agg, w, t
Kaputn.0	Average capacity utilization, change pres vs. prev q	diff, t
Kaputn.1	Average capacity utilization, change next vs. pres q	diff, t
PrisD.0	Industrial prices, home, change pres vs. prev q	diff, t
PrisD.1	Industrial prices, home, change next vs. pres q	diff, t
PrisF.0	Industrial prices, abroad, change pres vs. prev q	diff, t
PrisF.1	Industrial prices, abroad, change next vs. pres q	diff, t
OrdreD.0	Influx of orders, home marked, change pres vs. prev q	diff, t
OrdreD.1	Influx of orders, home marked, change next vs. pres q	diff, t
OrdreF.0	Influx of orders, export marked, change pres vs. prev q	diff, t
OrdreF.1	Influx of orders, export marked, change next vs. pres q	diff, t
Generell.1	General industry uncertainty judgment, next q	diff, t
Ressurs	Industrial indicator of resource scarcity, q	diff, t
Konj.EU	Compound industrial indicator, q	diff, t
Lag.com	Industrial storage of primaries at end of each q	diff, t
Lag.pr	Storage of retail products at end of each q	diff, t
L.c.pr	Industrial storage of primary vs. produced goods at end of each q	diff, t
L.v.o	Industrial storage of self produced vs. turnover, at end of each q	diff, t
Invest	Change in industrial investment plans, at end of each q	diff, t
Ord.Prod	Industrial orders vs. production, at end of each q	diff, t
SKI.s	Compound industry indicator, leading for production, q	s, t

Table C.1: Manufacturing indicators

Series name	Explanation	Transformations
Lager	Total for consumer good industry, volume index, change, q	s
OTi	Total for consumer good industry, volume index, supply, change, q	t
ORi	Total for consumer good industry, volume index, reserve, change, q	t
AIi	Investment index for energy (water, bio, etc), change q	s
K2real	Domestic credit to general public, change, q	s
K2hus	Domestic credit to households, change, q	s

Table C.2: Other indicators

Series name	Explanation	Transformations
ORtot	Compound ind. total of construction and factories, q	s
ORbol	Order reserve, value index, housing, q	s
OTbol	Order supply, value index, housing, q	s
ORabygg	Order reserve, value index, other construction, q	s
OTabygg	Order supply, value index, other construction, q	s
ORanlegg	Order reserve, value index, factories, q	n
OTanlegg	Order supply, value index, factories, q	n, t
OTba	Order supply, total of construction and factories, q	t
OTbygg	Order supply, total of construction, q	t
ORbygg	Order reserve, total of construction, q	t

Table C.3: Order reserves and supply, construction and factories

Series name	Explanation	Transformations
behning	Vacant positions reserve, change, m	s
tilgang	Vacant positions supply, change, m	s
ukeverk	Working hours pr. week, change , q	s, t
syss	Total employment, change, break, q	s
svare	Employment, construction and factories, change, break, q	s
sfintj	Employment, transport and communication, change, break, q	s
stjen	Employment, service sector, change, break, q	s
sindu	Employment, financial services, change, break, q	s
stransp	Employment, industry, change, break, q	s
sba	Employment, other services, change, break, q	s

Table C.4: Employment

Series name	Explanation	Transformations
Indikator	Compound index, average, change, q	s
Store-s	Normal goods, change, q	s, p
Landet-0-s	Norwegian economy last year, change, q	s
Landet-1	Norwegian economy next year, change, q	n
Egen-0	Household economy last year, change, q	n
Egen-1	Household economy next year, change, q	n

Table C.5: Norwegian trend indicators

Appendix D

Best models

These tables contain the 10 best individual *ex post* models for the first, second and third forecast horizon. Models that are highlighted in *italic* writing are three variable VAR models. The rest are four variable VAR models. These models are out-of-sample based forecasts made by the benchmark models described by the "Benchmark models" subsection in chapter 2. The RMSEs from bivariate VAR models were considerably larger than those of the best models presented here. Because of the lack of time the 4th and 5th step were not run since the first steps are the most important.

Rank	Model_name	Bias	RMSE	Forecast	Lags	Window
1	SKI.s OrdreF.1 stjen	-0.05	0.69	0.52	3	Recursive
2	Kaputn.1 Kaputn N.1	-0.01	0.70	-0.88	4	Recursive
3	SKI.s OrdreF.1 L.v.o	0.02	0.70	-0.14	3	Recursive
4	SKI.s N.1 OrdreF.1	-0.03	0.70	-0.16	4	Rolling
5	Konj.EU OrdreF.1 stjen	-0.07	0.72	-0.24	3	Recursive
6	*OrdreF.1 SKI.s*	-0.04	0.73	0.27	3	*Recursive*
7	SKI.s OrderF.0 stjen	0.05	0.73	1.13	3	Recursive
8	SKI.s Invest stjen	0.02	0.74	-0.18	3	Rolling
9	SKI.s OrdreF.1 sfintj	-0.04	0.74	0.49	3	Recursive
10	Kaputn SKI.s OrdreF.1	-0.05	0.74	-0.14	3	Recursive

Table D.1: Best ex post models for 1-step forecasting

Rank	Model_name	Bias	RMSE	Forecast	Lags	Window
1	Kaputn SKI.s Generell.1	-0.20	0.66	-1.91	4	Rolling
2	Kaputn SKI.s N.1	0.06	0.72	-1.62	3	Rolling
3	SKI.s N.1 OrdreF.1	-0.01	0.72	-0.89	3	Rolling
4	Kaputn SKI.s N.1	0.10	0.74	-1.76	4	Rolling
5	Kaputn N.1 Generell.1	0.01	0.75	-1.25	3	Rolling
6	Konj.EU N.1 OrdreF.1	-0.08	0.75	-2.61	3	Rolling
7	SKI.s N.1 OrdreF.1	-0.01	0.76	-0.96	4	Rolling
8	*Kaputn SKI.s*	-0.02	0.76	-1.44	4	*Rolling*
9	Kaputn N.1 OTbygg	-0.05	0.76	-1.09	4	Recursive
10	*N.1 Kaputn*	0.01	0.77	-1.79	4	*Rolling*

Table D.2: Best ex post models for 2-step forecasting

Rank	Model_name	Bias	RMSE	Forecast	Lags	Window
1	Kaputn SKI.s Generell.1	-0.13	0.82	-3.06	4	Rolling
2	Kaputn Konj.EU Generell.1	-0.24	0.87	-4.63	4	Rolling
3	SKI.s Generell.1 stjen	-0.04	0.93	-1.13	3	Rolling
4	Konj.EU Generell.1 stjen	-0.24	0.94	-3.69	4	Rolling
5	Konj.EU SKI.s Generell.1	-0.14	0.95	-5.09	4	Rolling
6	Konj.EU Generell.1 stjen	-0.23	0.95	-3.68	3	Rolling
7	Konj.EU L.v.o stjen	-0.07	0.95	-1.80	3	Rolling
8	OTbygg Invest sfintj	-0.15	0.95	0.68	4	Rolling
9	*Generell.1 Konj.EU*	-0.10	0.96	-3.87	4	*Rolling*
10	Konj.EU SKI.s Generell.1	-0.18	0.96	-4.64	3	Rolling

Table D.3: Best ex post models for 3-step forecasting

Appendix E

Program code

This section contains some examples and parts of models that were made for this project. Including all the models would use to much space, therefore only parts are included for illustrative purposes.

E.1 Code for the 4VAR, recursive window, MSE weighted model

```
library("vars");
library("xtable");
library("stats");

maxstepsize <- 5;

# defines max size of steps for the iterated  forecast

nvmax<- 12;

# number columns to be analyzed

#--------------------
## set value for lag order (used in VAR call) between 1 and 4

lagorder <- 4;

#--------------------

datatable <- read.csv2("vekt.csv", header=TRUE); #get input data
```

```
# biasAll[col2,col3,col4,stepsize] contains all bias
# calculations where col2 is a second column number
# from datatable, and col3 in a 3.column number from
# tabledata, col4 in a 4.column number from tabledata

biasAll <- c(NA);
length(biasAll) <-nvmax*nvmax*nvmax*nvmax*maxstepsize;
dim(biasAll)<- c(nvmax,nvmax,nvmax,nvmax,maxstepsize);

#---------------------------------------------------
# same for RMSE and forecast all
# (all forecasted values)

RMSEAll <- c(NA);
length(RMSEAll) <-nvmax*nvmax*nvmax*nvmax*maxstepsize;
dim(RMSEAll)<- c(nvmax,nvmax,nvmax,nvmax,maxstepsize);

#---------------------------------------------------

forecastAll <- c(NA);
length(forecastAll) <-nvmax*nvmax*nvmax*nvmax*maxstepsize;
dim(forecastAll)<- c(nvmax,nvmax,nvmax,nvmax,maxstepsize);

#---------------------------------------------------
# forecastAllknown[col2,col3,col4,maxstepsize,ind] contains all
# forecasts calculations where col2 is a column number
# from the datatable, and col3, col4 in a column number from the
# datatable, ind is defined by ind1=87 and ind2=121 for
# each stepsize from 1 to ind2-ind1+1

forecastAllknown <- c(NA);
length(forecastAllknown) <-nvmax*nvmax*nvmax*
        (121-87+1+maxstepsize)*maxstepsize;
dim(forecastAllknown)<- c(nvmax,nvmax,nvmax,maxstepsize,
        (121-87+1)+maxstepsize);

#---------------------------------------------------

# forecweighted[k,step] contains  equal weighted forecast
# for k from 1 to 121-87+1+maxstepsize for step=1,...,5

forecweighted<-c(NA);
```

```
length(forecweighted)<-(121-87+1+maxstepsize)*maxstepsize;
dim(forecweighted)<-c((121-87+1+maxstepsize),maxstepsize);

#---------------------------------------------------
# maxcolnumber defines number of columns from datatable
# that will be analyzed, normally its nvmax.

maxcolnumber<-nvmax;

#---------------------------------------------------


col2<-3;
while (col2<maxcolnumber-1){
col3<-col2+1;
        while (col3<maxcolnumber) {
        col4<-col3+1;
                while (col4<maxcolnumber+1) {
                msize <- max(length(datatable[[2]]),
                        length(datatable[[col2]]),length(datatable[[col3]]),
                        length(datatable[[col4]]) );
                PDSteps <- c(NA);
                length(PDSteps) <- (msize+maxstepsize)*maxstepsize*4;
                dim(PDSteps)<-c(maxstepsize, msize+maxstepsize, 4);

                startindex <-1;
                        while (is.na(datatable[[2]][startindex]) ||
                                is.na(datatable[[col2]][startindex])
                                        || is.na(datatable[[col3]][startindex])
                                        || is.na(datatable[[col4]][startindex]) ){
                                startindex<-startindex+1;
                        }

                y <-c(datatable[[2]], datatable[[col2]], datatable[[col3]],
                        datatable[[col4]]);
                dim(y)<-c(msize,4);

                index<-86;
                for(i in 1:36){
                        yy<-y;
                                # copy of y that will be itered.
                                # updated by forecasted values
                        dim(yy)<-c(msize,4);
```

```
                        for (stepsize in 1:maxstepsize){
                                newindex <- index+stepsize-1;
                                pp<-lagorder;
                                VARmodel <- VAR(yy[startindex:newindex,], p = pp,
                                        type = "none", season = NULL, exogen = NULL,
                                        lag.max = NULL);
                                VARforecast <- predict(VARmodel, n.ahead = 1,
                                        dumvar = NULL);
                                yy[index+stepsize,]
                                        <- c(VARforecast$fcst$y1[1],
                                        VARforecast$fcst$y2[1],
                                        VARforecast$fcst$y3[1],
                                        VARforecast$fcst$y4[1]);

                                        PDSteps[stepsize,index+stepsize,]
                                        <- c(VARforecast$fcst$y1[1],
                                        VARforecast$fcst$y2[1],
                                        VARforecast$fcst$y3[1],
                                        VARforecast$fcst$y4[1]);
                        }
                index <- index+1;
                }
#------------------------------------------------
#remember all forecasted values for model defined by col2, col3 and col4

                forecastAllknown[col2,col3,col4,,]
                        <-PDSteps[,87:(121+maxstepsize),1];

#------------------------------------------------

                col4<-col4+1;
                }
             col3<-col3+1;
        }
        col2<-col2+1
}

#----------------------------------------------
# Now we will calculate weighted forecasts
# by using forecastAllknown and MSE weights
#----------------------------------------------

for (step1 in 1:maxstepsize){
```

```
        r1<-forecastAllknown[,,,step1,step1];
        r2<-r1[!is.na(r1)];
        summa<-(y[86+step1,1]-r2)^2;

        for (k in 1:(35-step1)){
                r1<-forecastAllknown[,,,step1,(k+step1)];
                r2<-r1[!is.na(r1)];
                summa<-summa+(y[86+step1+k,1]-r2)^2;
                MSE<-summa/(k+1);
                ss<-sum(1/MSE);
                W<-1/(MSE*ss); #MSE weights
                r1<-forecastAllknown[,,,step1,(k+step1+step1)];
                r2<-r1[!is.na(r1)];
                r<-W*r2;
                forecweighted[k+step1+step1,step1]<-sum(r);
        }
}
#------------------------------------------------
# bias for weighted calculations for steps 1,2,...,5
                biasweighted <- c(NA); length(biasweighted)
                <-maxstepsize;
                ind1<-87;
                ind2<-121;      # upper border for forecast bias calc.
                for (j in 1:maxstepsize){
                        biasweighted[j]
                        <-mean(y[(ind1+2*j):ind2,1]
                        -forecweighted[(2*j+1):(ind2-86),j]);
                }
#-----------------------------------------------
# RMSE for weighted calculations
                RMSEweighted <- c(NA); length(RMSEweighted)
                <- maxstepsize;
                ind1<-87;
                ind2<-121;      # upper border for RMSE forecast calc.
                for (j in 1:maxstepsize){
                        RMSEweighted[j]
                        <- sqrt(mean((y[(ind1+2*j):ind2,1]
                        -forecweighted[(2*j+1):(ind2-86),j])^2));
                }

#----------------------------------------------------------
# Data for ensembles of forecast
FW4VARfixiterMSE <-forecweighted;
```

```
#----------------------------------------------------------

## table preparation:
#  resulttable[step] contains table for step step=1,2,...,5

        resulttable<-as.data.frame(NA); length(resulttable)
                <-maxstepsize;
        resulttable<-data.frame(Step=c(1:5),
                Bias=biasweighted, RMSE=RMSEweighted);

        tablename<-paste("Table for 4VAR MSE weighted fix-iter");
        filename<-paste("4VAR-recursive-iter-weightedMSE",".tex",sep="");

        LTtable<- xtable(resulttable, caption=tablename,
                label=NULL, align=NULL, digits=NULL);

        # filename below contains a name of file defining where
  # the table will be saved in latex format.
        print(LTtable, type="latex",
                include.rownames=FALSE, file=filename);

#-----------------------------------------------------------
# Prep data for time series plotting
# Creates a "hairy line graph"
#------------------------------------------------------------

gdp <-ts(datatable[[2]][86:121], frequency = 4,
        start = c(1999, 4), end=c(2010,1));
gdp[37:42]<-NA;

pdf(file=paste("4VAR-recursive-MSE-weigted-all-steps",".pdf",sep=""),
        paper="special", width=15, height=10)
ts.plot(gdp,col=c("black"),lwd=c(4),
        main=paste(" "),
        gpars=list(xlab="Years", ylab="GDP growth",
                type=c("o"),lty=c(1)));

vint<-0;
vhair<-c(datatable[[2]][86+vint],forecweighted[vint+1,1],
        forecweighted[vint+2,2],
        forecweighted[vint+3,3],
        forecweighted[vint+4,4],
        forecweighted[vint+5,5]);
```

```
hairyplot<-ts(vhair, frequency = 4, start = c(1999,4));
lines(hairplot,col=c("red"),type="l");
vint<-vint+1;

for (year in 2000:2007){
        for (quater in 1:4){
                vhair<-c(datatable[[2]][86+vint],forecweighted[vint+1,1],
                        forecweighted[vint+2,2],
                        forecweighted[vint+3,3],
                        forecweighted[vint+4,4],
                        forecweighted[vint+5,5]);
                hairplot<-ts(vhair, frequency = 4, start = c(year,quater));
                lines(hairplot,col=c("red"),type="l");
                vint<-vint+1;
        }
}
for (quater in 1:3){
                vhair<-c(datatable[[2]][86+vint],forecweighted[vint+1,1],
                        forecweighted[vint+2,2],
                        forecweighted[vint+3,3],
                        forecweighted[vint+4,4],
                        forecweighted[vint+5,5]);
                hairplot<-ts(vhair, frequency = 4, start = c(2008,quater));
                lines(hairplot,col=c("red"),type="l");
                vint<-vint+1;
        }

dev.off()
```

E.2 Program code; AIC weights for 4VARs

```
n4var<-4;
T<- 63-lagorder;
VC<-summary(VARmodel)$covres; # extract varcov matrix from VARmodel
VCold<-VC;
invVC<-solve(VC);
DOmega<-det(invVC);
Lambda<- 0.5*T*(log(DOmega)-n4var*log(2*pi)-n4var);
K<-n4var*(n4var*pp + 1) + n4var^2;
AICvalues[stepsize,index+stepsize]<-(2*K-2*Lambda);

#----------------------------------------------
# Calculating weighted forecast by using
# forecastAllknown and AIC values.
#----------------------------------------------

for (step1 in 1:maxstepsize){

        for (k in 1:36){
                rr<-AICvaluesAllknown[,,,step1,(k+step1-1)];
                delta[,,,step1,(k+step1-1)]
                        <-AICvaluesAllknown[,,,step1,(k+step1-1)];
                re[,,,step1,(k+step1-1)]
                        <-exp(-0.5*delta[,,,step1,(k+step1-1)]);
                SW[step1,(k+step1-1)]
                        <-sum(re[,,,step1,(k+step1-1)]
                        [!is.na(re[,,,step1,(k+step1-1)])]);
                W[,,,step1,(k+step1-1)]
                        <-re[,,,step1,(k+step1-1)]/SW[step1,(k+step1-1)];
                rf<-forecastAllknown[,,,step1,(k+step1-1)]
                        *W[,,,step1,(k+step1-1)];
                forecweighted[(k+step1-1),step1]<-sum(rf[!is.na(rf)]);
        }
}
```